Identification of novel tumor associated proteins in head and neck cancer using patient derived fresh tissue cDNA libraries and antibody repertoires

Von der Fakultät für Lebenswissenschaften

der Technischen Universität Carolo-Wilhelmina zu Braunschweig

zur Erlangung des Grades eines

Doktors der Naturwissenschaften

(Dr. rer. nat.)

genehmigte

D i s s e r t a t i o n

von Kilian Johannes Carl Zilkens

aus Köln

1. Referent: Prof. Dr. Stefan Dübel
2. Referent: Prof. Dr. Andreas Gerstner
eingereicht am: 10.02.2020
mündliche Prüfung (Disputation) am: 19.06.2020

Druckjahr 2020

Veröffentlichungen der Dissertation

Teilergebnisse aus dieser Arbeit wurden mit Genehmigung der Fakultät für Lebenswissenschaften, vertreten durch den Mentor der Arbeit, in folgenden Beiträgen vorab veröffentlicht:

Posterbeiträge:

Zilkens, K., Philippi, M., Kügler, J., Sellmann, C., Clarke, T., Toleikis, L., Dübel, S., Schirrmann, T.: Antibody and target discovery using tumor-infiltrating B lymphocytes. Poster 72. 12th PhD Symposium der HZI International Graduate School for Infection Research (GS-FIRE) am Helmholtz Zentrum für Infektionsforschung, Braunschweig (2019)

Abstract

As a result of the advancing development of technologies and applications in immunotherapy there is an increasing demand for the discovery of novel therapeutic targets. One of the main drivers in this progression is cancer research, which is manifested in 2,305 promising immunotherapeutic drugs directed against 378 targets currently in preclinical development. Novel applications of immunotherapies in turn require specific biomarkers to guide treatment decisions, monitor disease progression and to predict patient outcome, further increasing the need for discovery. Typical discovery ventures make use of proteomics and transcriptomics approaches that try to provide solutions via big data or rely on screening of common peptide or protein libraries against large numbers of patient sera. Though effective in some cases, such large scope techniques generally do not take into account patient-specific phenotypes and might thereby miss individual indications.

This study presents a novel approach for the selection of potential biomarkers and therapeutic targets using fresh untreated tumor tissue and patient antibody repertoires. Patient specific cDNA antigen libraries were created from head and neck cancer tissues. Autologous antibody repertoires obtained from tumor infiltrating B cells (TIBC) or patient sera were used for selection of protein fragments via M13 phage display.

The technology was tested successfully by selecting mesothelin (MSLN) peptides from HCT116 cell line antigen libraries using anti-MSLN antibody SUW57-D11. Panning in solution in combination with phage ELISA screening for positive clones was found to be the most promising workflow for biomarker and target discovery. High throughput applicability of this method was confirmed through successful selections in 96-well format and by employing next generation sequencing (NGS) as a single massively parallel screening method to replace ELISA.

Using antibodies from TIBC, protein fragments of known cancer therapy target MMP9 were isolated from the antigen libraries, providing proof of principle for the identification of potential therapeutic targets. Furthermore, SHTN1 and WWC2 protein fragments were enriched in a similar approach. Using patient sera, EEA1 emerged as a protein of interest.

The presented technology allows target and biomarker discovery based on patient specific local and systemic immune responses. This approach yields candidates with an implied relevance by basing the selection process on a biological mechanism within the disease context. The workflow has the capacity to be advanced towards a target and biomarker discovery platform that is applicable not only to cancer, but to diseases that require immunotherapeutic interventions in general.

Table of contents

List of Tables

List of Figures

List of Abbreviations

°C	Degree Celsius
µl	microliter
µm	micrometer
2YT	Two times yeast tryptone
aa	amino acid
ACT	Adoptive cell therapy
ACTB	Beta-actin
ADCC	Antibody dependent cell mediated cytotoxicity
AK	Supplemented with ampicillin and kanamycin
ATCC	American Type Culture Collection
AT-IPTG	Supplemented with ampicillin, tetracycline and IPTG
BLAST	Basic local alignment search tool
BLASTn	BLAST for nucleotide sequences
bp	base pairs
CAR	Chimeric antigen receptor
CD	Cluster of differentiation
CDC	Complement dependent cytotoxicity
cDNA	Complementary DNA
CEACAM5	CEA Cell Adhesion Molecule 5
CES1	Carboxylesterase 1
CHD8	Chromodomain Helicase DNA Binding Protein 8
CIP	Calf intestinal phosphatase
cm	centimetre
ctrl	control
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
dNTPs	Deoxynucleotide triphosphates
EDTA	Ethylenediaminetetraacetic acid
EEA1	Early Endosome Antigen
EFHD2	EF-Hand Domain Family Member D2

EGFR	Epidermal Growth Factor Receptor
ELISA	Enzyme-linked immunosorbent assay
FBS	Fetal bovine serum
Fc	Fragment crystallizable
FDA	Food and Drug Administration
GA	Supplemented with glucose and ampicillin
GAPDH	Glyceraldehyde-3-Phosphate Dehydrogenase
GAT	Supplemented with glucose, ampicillin and tetracycline
h	hour
HEK	Human embryonic kidney
HRP	Horseradish peroxidase
HSPA8	Heat Shock Protein Family A (Hsp70) Member 8
IgG	Immunoglobulin G (gamma (γ) heavy chain)
IGHV	Immunoglobulin heavy chain variable domain
IPTG	Isopropyl-β-D-thiogalactopyranosid
kB	kilobases
kDa	kilodalton
kV	kilovolt
l	liter
LB	Lysogeny broth
M	molar
mg	milligram
min	minute
ml	milliliter
mM	millimolar
MMP9	Matrix metallopeptidase 9
mRNA	Messenger RNA
MSLN	Mesothelin
MTCO2P12	Mitochondrially encoded cytochrome C oxidase II pseudogene 12
MTP	microtiter plate
MYC	MYC Proto-Oncogene
N	Normality

NCBI	National Center for Biotechnology Information
ng	nanogram
NGS	Next generation sequencing
NK cells	Natural killer cells
nm	nanometer
nM	nanomolar
OD600	Optical density at 600 nm
ORF	Open reading frame
PBS	Phosphate buffered saline
PBST	Phosphate buffered saline supplemented with Tween20
PCR	Polymerase chain reaction
pIII	M13 phage minor coat protein 3
PRKDC	Protein Kinase, DNA-Activated, Catalytic Subunit
PRRC2B	Proline Rich Coiled-Coil 2B
pVI	M13 phage minor coat protein 6
pVIII	M13 phage major coat protein 8
RNA	Ribonucleic acid
rpm	rounds per minute
s	second
scFv	Single chain fragment variable
SEREX	Serological identification of antigens by recombinant expression cloning
SHTN1	Shootin 1
SOB	Super optimal broth
SOC	Super optimal broth with catabolite repression
SPRI	Solid-phase reversible immobilization
SPRR3	Small Proline Rich Protein 3
T	Supplemented with tetracycline
TAE	TRIS-Acetate-EDTA
TIBC	Tumor-infiltrating B cells
TMB	Tetramethylbenzidine
TOP1MT	DNA Topoisomerase I Mitochondrial
TU	Technische Universität

U	unit
UBE2L6	Ubiquitin Conjugating Enzyme E2 L6
UCSC	University of California, Santa Cruz
UV	Ultraviolet
V	Volt
v/v	Volume per volume
w/v	Weight per volume
WWC2	WW Domain-Containing Protein 2

1 Introduction

1.1 Disease background

1.1.1 Cancer biology

Cancer is a disease that arises due to a continuous and progressing dysregulation of various cellular functions. With few exceptions (for example in childhood cancers) these are originally caused by gradual accumulation of DNA damage and mutations that change gene regulation and protein expression and function. Over time, a complex interplay of different cell types and signalling pathways develops, that enables advancement of the disease. The most fundamental change in tumor cells is their ability to proliferate indefinitely, achieved through sustained activation of proliferative signalling pathways and suppression of growth inhibitors. This constant growth requires a constant supply of nutrients which growing tumors establish via induction of angiogenesis, the formation of new blood vessels. During further expansion, invasion of surrounding tissues is facilitated through manipulation of the extracellular matrix leading to a reduction of local tissue cohesion and an increase in cell motility. In later stages some tumor cells undergo epithelial-mesenchymal transition allowing complete dissemination from the primary tumor and invasion of distant organs via lymphatic and blood vessels[1].

Besides these intrinsic hallmark capabilities, tumors are also influenced by their surrounding microenvironment. Next to cells of blood vessels and fibroblasts many cells of the immune system, for example macrophages, NK cells and T and B lymphocytes, are found within or in the vicinity of solid tumors[2]. Increased local immune signalling can cause tumors to become sites of chronic inflammation leading to further tumor-promoting effects, for example release of mutagenic reactive oxygen species[1]. While T and B cells have been shown to elicit tumor specific immune responses, the execution of immune effector functions is often impaired, in part due to suppressive signals in the microenvironment[3]. These various genetic and molecular changes coupled with a complex interaction network involving different cell types make cancer a highly heterogenous disease the treatment of which requires precise diagnostics and highly specific targeted therapies.

1.1.2 Tumor infiltrating B cells

Solid tumors are frequently infiltrated by lymphocytes[4,5]. Tumor-infiltrating lymphocytes have strong prognostic potential in many cancers, yet depending on the type of disease and the subtype of cells they can have either beneficial or detrimental effects[6]. In general, the effects of infiltrating T cells have been studied to a greater extent than B cells, due to their direct cytotoxic capabilities, but there is evidence that tumor infiltrating B cells (TIBC)

do elicit a humoral immune response and that their antibody repertoires may be used to target cancer cells[7]. Furthermore, tertiary lymphoid structures can form in solid cancers, which harbour B cells, T cells, dendritic cells and others[8]. Within these structures immune cells are constantly exposed to all sorts of antigens in the surrounding microenvironment. They can also allow somatic hypermutation and affinity maturation to take place which are important processes for the generation of highly specific antibodies[9]. This study uses antibody repertoires from TIBC from human head and neck cancer tissues for the identification of potential novel therapeutic targets.

1.1.3 Head and neck cancer

Head and neck cancer is a collective term for all cancers that arise from tissue of the lips, oral cavity, throat and upper airways. Incidences are especially high for males with 511,060 newly diagnosed patients in 2018 compared to 199,117 females[10]. Primary treatment for tumors classified T1-T3 is still surgery in combination with optional radiotherapy. Higher stage tumors are treated with primary radiochemotherapy[11]. Until recently, immunotherapies for head and neck cancer, for example with anti-EGFR antibody cetuximab, were sparse and mostly regarded as second-line treatments, but their development is on the rise and several immune checkpoint inhibitors have been approved, especially for recurrent and metastatic disease[12].

1.2 Cancer therapy development

1.2.1 Immune-oncology and immunotherapy

Cancer therapy is progressively moving away from unspecific chemotherapy towards targeted therapies[13]. This greatly reduces side effects due to unselective killing of normal cells and increases efficacy. Immune-oncology deals with the development and investigation of cancer treatments that use the mechanisms of the body's immune system to fight the disease. Despite it being a rapidly growing field of research only 25 protein targets can currently be addressed with approved immune-oncologic therapies, with 43 out 59 agents acting on only 9 targets[14,15]. There is a great need for the discovery and validation of new targets and the according therapies as demonstrated by 3817 agents directed against 466 targets currently in preclinical and clinical phase I-III trials[15].

Immunotherapies aim to activate the patients' immune system. They either target proteins or certain types of cells directly by provoking an immune response directed specifically against these structures or indirectly by eliciting a general immune response that attacks the target structure by coincidence. Cancer immunotherapies can be divided into three major approaches: (1) vaccination therapy, (2) antibody therapy and (3) adoptive cell therapy[16].

Vaccination therapy relies on injection of cancer cells, parts of cancer cells, purified proteins or protein fragments to cause a direct immune response together with adjuvants to boost the reaction. For therapy to be effective, the administered proteins need to be either highly overexpressed on cancer cells, a differentiation marker specific to the cancer cell type or a mutated neoantigen. Otherwise the vaccination has a high likelihood of failure due to self-tolerance mechanisms[17]. New sophisticated techniques use autologous antigen presenting cells (for example dendritic cells) that have been primed with antigen *ex vivo* as vaccination agents. The difficulty of creating effective cancer vaccines is demonstrated by the fact that Spileucel-T is the only FDA approved vaccination treatment against cancer (hormone-refractory prostate cancer) to date.

In antibody therapy, recombinant immunoglobulins directed against cancer antigens are used to target cancer cells. To reduce side effects, the targeted cancer antigens should have similar properties as described for vaccination therapy. Antibodies can cause tumor cell killing through immune-mediated mechanisms like complement dependent cytotoxicity (CDC), antibody dependent cell mediated cytotoxicity (ADCC) or regulation of T cell function. Independent of the immune system antibodies can affect tumors through inhibition or activation of signalling pathways resulting for example in induction of apoptosis. Antibodies that are rapidly internalized after binding their target can be used to deliver cytotoxic agents into the cell[18]. Bispecific antibodies can be employed to engage two targets at once for increased efficacy or to guide immune cells, for example cytotoxic T-cells, to the tumor[19]. Antibody therapeutics may also target the tumor microenvironment or stroma. Recently, immune checkpoint inhibitors have shown great promise in fighting cancer by releasing immune system blockades[20]. In addition to these versatile applications, the molecular structure of antibodies can be readily modified to improve their half-life, solubility and stability which makes them a very flexible tool for immunotherapy.

Adoptive cell therapy (ACT) makes use of *ex vivo* expanded and/or modified autologous lymphocytes. Tumor infiltrating lymphocytes with the potential to exert anti-tumor activities are harvested by biopsies and cultivated to great numbers. After eradication of other lymphocytes in the patient's body through radiation treatment, the expanded cells are reinjected to fight the cancer. Via genetic engineering, peripheral blood T cells can be modified to express highly antigen specific T cell receptors to recognize cancer cells. However, T cell receptor can only recognize short peptides presented via MHC complexes. This was overcome by the development of chimeric antigen receptor T (CAR-T) cells. T cell receptors are endowed with the antigen binding region of an antibody vastly increasing the possible target space for adoptive cell therapies. Many ACT are still at an

early stage of development but clinical evidence has shown promising efficacy in the treatment of cancers[21].

1.2.2 Target discovery

The development of a drug is preceded by the search for a molecular target that is suitable for therapeutic intervention. The target should ideally have a function in disease biology, be druggable and not uniformly expressed in the body[22]. Here, several approaches to select protein target candidates for further investigation of their therapeutic potential are described.

Targets can be identified via fundamental research and elucidation of disease biology leading to the discovery of novel molecules and proteins that participate in a disease mechanism. In a rational approach using literature research and data mining key players in disease development can be identified and their potential for therapeutic interventions be explored. This process requires great manual effort and has a high risk of failure in the following drug development process[23]. However, the advantage of selecting targets through deduction from large amounts of data lies within their advanced molecular characterization.

Target discovery can also be built on previous successful drug discoveries. For example, through investigation of members of protein families that have proven to be more readily druggable then others[24]. Similarly, members of effective compound classes that are not yet indicated for therapeutic use can be tested in disease models. High throughout screenings provide the capacity to test thousands of compounds, for example on mammalian cancer cell lines. Effective drug candidates are selected via measurement of phenotypical reactions and the molecular target is later identified via target deconvolution procedures[25].

Recently, omics technologies (genomics, proteomics, transcriptomics, metabolomics) that aim to gather data on all DNA, protein, mRNA or metabolite molecules in a biological sample have been employed. These approaches aim to provide a wholistic view of all factors that may have implications in disease. However, they are not driven by a specific biological mechanism. Factors that are altered in a disease context are identified through creation of large databases and comparative analyses[26].

Another target discovery approach relies on exploiting biological mechanisms, for example the host immune response. The immune system is a refined biological mechanism for targeting biological structures. Patient antibodies can be used to identify novel relevant proteins via microarrays[27], mass spectrometry[28] or phage display[29]. This approach is

especially interesting for identifying targets for immunotherapies which are often guided by antibody binding specificities. Using recombinant technologies, the potential therapeutic target and corresponding antibodies can be selected in parallel.

1.2.3 Biomarkers

Biomarkers can generally be defined as “a biological observation that substitutes for and ideally predicts a clinically relevant endpoint or intermediate outcome that is more difficult to observe”[30]. A typical biomarker in cancer would be the expression level of a protein that helps to guide precise diagnosis, tumor subtyping or choice of treatment or that can be predictive of disease progression. Biomarkers are especially important in cancer treatment, since it is a very heterogenous disease that progresses over very long periods of time, sometime decades. The ability to make evidence-based treatment decisions as early as possible is crucial for the patients’ well-being and survival.

1.3 Phage and ORFeome display

In general, phage display is used to present proteins or peptides of interest on the surface of bacteriophages, for example T7 or M13, to study protein-protein interactions. This is achieved by fusing the proteins to one of the coat proteins of the phage. The many different phage display systems and applications have been reviewed previously[31]. The system that is the basis for this study accomplishes display by fusion of the desired construct to the minor coat protein pIII of bacteriophage M13[32]. pIII is responsible for binding of M13 to the F-pilus of bacteria and thereby crucial for infection. The genetic Information for the protein-pIII fusion is encoded on a phagemid vector which can be packaged into the phage capsid and be propagated in bacteria, most commonly *Escherichia coli*.

The process of selection of protein interaction partners in phage display is called panning and consists of four major steps:

1) immobilization of an interaction partner on a surface
2) addition of a library of proteins/peptides, displayed on the phage surface, that potentially bind the interaction partner
3) removal of unbound phage
4) collection of specifically bound phage.

In a variation, the interaction partner can be applied in solution and captured on a surface, e.g. magnetic beads, at the end of the selection process (Figure 1-1). The selected phage can be amplified by infection of *E. coli* and used for further rounds of selection and enrichment of specific binders. This approach has been very successful in generating

specific antibodies to various antigens[33,34] and has also been used to identify epitopes of antibodies in a reverse manner[35]. Effectiveness of polypeptide presentation was improved by introduction of the pHORF system, also called ORFeome display. First developed in 2006[36], it has so far been used to identify immunogenic peptides and proteins from bacteria[37–39], metagenomes of gut microbiota[40] and tick saliva[41]. It uses Hyperphage[42] to enrich open reading frames (ORF) in polypeptide libraries constructed from non-directionally cloned, randomly fragmented cDNA or genomic DNA. The pIII protein has been deleted from the Hyperphage genome so the only source of pIII are intact i.e. in frame pIII-fusion constructs which causes their enrichment upon Hyperphage packaging. Without ORF enrichment, only 5.55 % of cloned peptide constructs would be in frame with pIII and displayed on the phage surface, which greatly reduces efficiency. In this study, the ORFeome display technology was improved for the use with human cDNA libraries by implementing a more efficient unidirectional cloning process that by design doubles the statistical occurrence of in frame constructs prior to Hyperphage packaging.

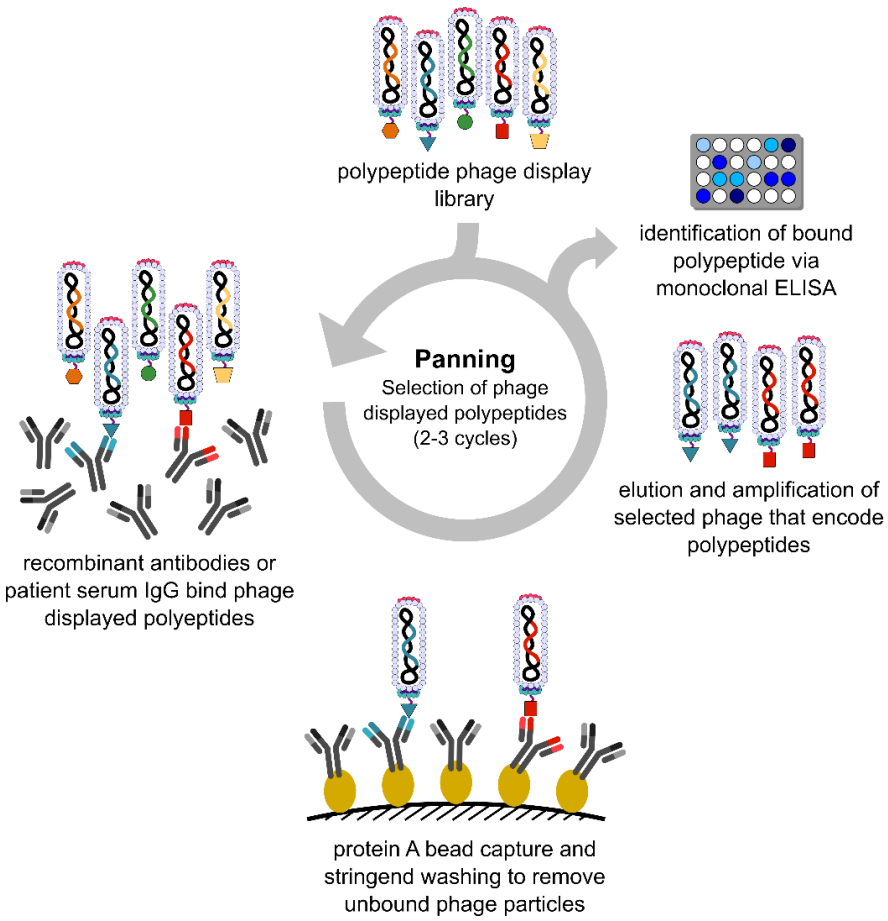


Figure 1-1 – Panning in solution: *Schematic overview of the panning procedure. Libraries consisting of phage particles that are displaying polypeptides on their surface are mixed with antibodies. Antibodies bind to polypeptides according to their specificity. Protein A magnetic beads are used to capture antibodies and unbound phage particles are washed away. Bound phage particles are eluted and either amplified for another round of selection or used for single clone screening ELISA. (adapted from Kügler et al.[43])*

1.3.1 Objectives

This study aims to develop a technology that can select protein fragments that have the potential to become new targets or biomarkers for cancer therapy. Using recombinant antibody repertoires from TIBC isolated from fresh head and neck tumor tissue and using patient sera, protein fragments are enriched from antigen phage libraries via panning. Furthermore, high throughput applicability of the approach is investigated that may allow this technology to be developed into a discovery platform.

2 Materials and Methods

2.1 Materials

2.1.1 Consumables

Consumables used in this study are listed in Table 2-1.

Table 2-1: *Consumables*

Material	Manufacturer
Amicon Ultra-0.5 ml Centrifugal Filters, 30K	Merck KGaA (Darmstadt, DE)
Cell strainer, 40 µm	Fisher Scientific GmbH (Schwerte, DE)
Costar 96 well plate	Corning Inc. (Corning, NY, USA)
Cryo.S Biobankröhrchen, 1000 µl with 2D codes	Greiner-Bio-One (Frickenhausen, DE)
Cryo.S Biobankröhrchen, 300 µl with 2D codes	Greiner-Bio-One (Frickenhausen, DE)
Deepwell plate, 24-well, Whatman Uniplate, PP, 10 ml	GE Healthcare (Chicago, IL, USA)
Deepwell plate, 96-well, Riplate® medio, PP, 0,5 ml	Ritter GmbH (Schwabmünchen, DE)
Disposable Cuvettes, PS, 10 mm	Sarstedt AG & Co. KG (Nümbrecht, DE)
Drigalski spatula, sterile	VWR International GmbH (Hannover, DE)
Falcon centrifuge tubes, 15 ml	Corning Inc. (Corning, NY, USA)
Falcon centrifuge tubes, 50 ml	Corning Inc. (Corning, NY, USA)
GenePulser Cuvette, 0.1 cm	Bio-Rad Laboratories (München, DE)
Minisart sterile filter, 0.45 µm	Sartorius AG (Göttingen, DE)
MTP 384 well PS-F-bottom	Greiner-Bio-One (Frickenhausen, DE)
MTP 96 well PP F-bottom	Greiner-Bio-One (Frickenhausen, DE)
MTP cover foil, aluminium	HJ Bioanalytik GmbH (Erkelenz, DE)
MTP cover foil, nonwoven	HJ Bioanalytik GmbH (Erkelenz, DE)
Multiply-µStrip Pro8-strip PCR stripes	Sarstedt AG & Co. KG (Nümbrecht, DE)
Nitrile gloves, StarGuard	STARLAB INTERNATIONAL GmbH (Hamburg, DE)
Nunc Bio Assay Dish (pizza plate)	Merck KGaA (Darmstadt, DE)
PCR plate 96 well, frameless	Sarstedt AG & Co. KG (Nümbrecht, DE)
PCR plate 96 well, Semi-Skirted	4titude Ltd (Berlin, DE)
PCR seal	4titude Ltd (Berlin, DE)
Petri Dish, 10 cm	Greiner-Bio-One (Frickenhausen, DE)
Pipette tips, 125 µl	Integra Biosciences GmbH (Biebertal, DE)
Pipette tips, Combitips advanced: 5 ml, 10 ml, 25 ml, 50 ml	Eppendorf AG (Hamburg, DE)
Pipette tips, filtered: 10 µl, 20 µl, 300 µl, 1000 µl	Nerbe Plus GmbH (Winsen/Luhe, DE)

Material	Manufacturer
Pipette tips: 10 µl, 20 µl, 200 µl, 1000 µl	Sarstedt AG & Co. KG (Nümbrecht, DE)
Pipette tips: 300 µl V96 Tips NS & SF	Integra Biosciences GmbH (Biebertal, DE)
Reaction tubes: 1.5 ml, 2 ml	Sarstedt AG & Co. KG (Nümbrecht, DE)
Reservoir 300 ml, sterile	Integra Biosciences GmbH (Biebertal, DE)
Screw-top tubes, 2ml, PP	Sarstedt AG & Co. KG (Nümbrecht, DE)
Serological Pipettes: 2 ml, 5 ml, 10 ml, 25 ml	Corning Inc. (Corning, NY, USA)
SPRI select magnetic beads	Beckman Coulter GmbH (Krefeld, DE)
Strips of 8 Flat Optical Caps	4titude Ltd (Berlin, DE)
SureBeads™ Protein A Magnetic Beads	Bio-Rad Laboratories (München, DE)
Syringes: 2 ml, 10 ml, 20 ml	B.Braun Melsungen AG (Melsungen, DE)
TC-Flask T75, standard	Sarstedt AG & Co. KG (Nümbrecht, DE)

2.1.2 Technical devices

Technical devices used in this study are listed in Table 2-2.

Table 2-2: *Technical devices*

Device	Model	Manufacturer
Agarose Gel Combs	6, 10, 12, 50 wells	Peqlab Biotechnologie GmbH (Erlangen, DE)
Autoclave	VX-150	Systec GmbH (Linden, DE)
	VX-95	Systec GmbH (Linden, DE)
Balance	Entris 4202l-1S	Sartorius AG (Göttingen, DE)
	Genius ME215P	Sartorius AG (Göttingen, DE)
Centrifuge	Allegra X-15R	Beckman Coulter GmbH (Krefeld, DE)
	5810R	Eppendorf AG (Hamburg, DE)
	Sorvall LYNX 4000 Superspeed	Thermo Fisher Scientific (Dreieich, DE)
	Heraeus Pico 17 Microcentrifuge	Thermo Fisher Scientific (Dreieich, DE)
Colony Picker	Qpix	Molecular Devices, LLC (San Jose, CA, USA)
Counting chamber	BLAUBRAND® Neubauer, improved	BRAND GMBH + CO KG (Wertheim, DE)
Electrophoresis chamber	40-0708	Peqlab Biotechnologie GmbH (Erlangen, DE)
	40-2314-N	Peqlab Biotechnologie GmbH (Erlangen, DE)
Electroporation device	MicroPulser™	Bio-Rad Laboratories (München, DE)

Device	Model	Manufacturer
Freezer -20 °C		Liebherr-International Deutschland GmbH (Biberach an der Riß, DE)
Freezer -80 °C	Innova U725	Eppendorf AG (Hamburg, DE)
Fridge 4 °C		Liebherr-International Deutschland GmbH (Biberach an der Riß, DE)
Heating plate / magnetic stirrer	MR3001	Heidolph Instruments GmbH & Co. KG (Schwabach, DE)
Imaging System	ChemiDoc™	Bio-Rad Laboratories (München, DE)
Incubator	IS-2-K	Axon Labortechnik GmbH (Kaiserslautern, DE)
	Vortemp 56	Labnet International, Inc (Edison, NJ, USA)
	ZWYC-290A	LABWIT Scientific Pty. Ltd (Melbourne, AUS)
	HERAcell Vios 160i	Thermo Fisher Scientific (Dreieich, DE)
	Heraeus UT 6200	Thermo Fisher Scientific (Dreieich, DE)
Laminar Flow Hood	Heraguard ECO	Thermo Fisher Scientific (Dreieich, DE)
	Safe 2020	Thermo Fisher Scientific (Dreieich, DE)
Light Microscope	IX70	Olympus Europa SE & CO. KG (Hamburg, DE)
Magnetic Rack	96S Super Magnet Plate	Alpaqua Engineering, LLC (Beverly, MA, USA)
	DynaMag™-2	Thermo Fisher Scientific (Dreieich, DE)
Microplate plate washer	EL406	BioTek (Bad Friedrichshall, DE)
	ELx405	BioTek (Bad Friedrichshall, DE)
	Hydro flex	Tecan Group AG (Männedorf, CHE)
Microplate reader	Sunrise	Tecan Group AG (Männedorf, CHE)
Microplate Sealer	Quick-Combi Sealer Plus	HJ Bioanalytik GmbH (Erkelenz, DE)
Microplate Stacker	BioStack 3 & 4	BioTek (Bad Friedrichshall, DE)
pH Meter	WTW innoLab pH7110	Xylem Analytics Germany Sales GmbH & Co. KG (Weilheim, DE)
Photometer	ScanDrop^2	Analytik Jena (Jena, DE)
	BioPhotometer	Eppendorf AG (Hamburg, DE)
Pipette	Multipette M4	Eppendorf AG (Hamburg, DE)
Pipette electric 8-channel	Voyager	Integra Biosciences GmbH (Biebertal, DE)
Pipettes	Research Plus: 0,1-2,5 µl; 0,5-10 µl; 2-20 µl; 20-	Eppendorf AG (Hamburg, DE)

Device	Model	Manufacturer
	200 µl; 100-1000 µl	
Pipettes 8-channel	Research plus: 0,5-10 µl; 10-100 µl; 30-300 µl	Eppendorf AG (Hamburg, DE)
Pipettes electric 96-channel	Viaflo96	Integra Biosciences GmbH (Biebertal, DE)
Pipettor	Acu-Jet	Brand (Wertheim, DE)
Power supply	Power Pac HC 200 & 300	Bio-Rad Laboratories (München, DE)
Rotator	Multi Bio RS-24	SIA Biosan (Riga, LV)
Sample Rocker	Duomax 1030	Heidolph Instruments GmbH & Co. KG (Schwabach, DE)
Sample Roller	RS-TR05	Phoenix Instrument (Garbsen, DE)
Shaking plate	Unimax 2010	Heidolph Instruments GmbH & Co. KG (Schwabach, DE)
Thermal Cycler	T100	Bio-Rad Laboratories (München, DE)
	Mastercycler personal	Eppendorf AG (Hamburg, DE)
Thermal Mixer	Thermomixer comfort	Eppendorf AG (Hamburg, DE)
	13-687-717	Thermo Fisher Scientific (Dreieich, DE)
Water Purification System	Milli-Q UF Plus	Merck KGaA (Darmstadt, DE)
Water bath	Shake Temp SW23	Julabo GmbH (Seelbach, DE)

2.1.3 Chemicals

Chemicals and reagents used in this study are listed in Table 2-3.

Table 2-3: *Chemicals and reagents*

Chemical / Reagent	Supplier
2-propanol	Carl Roth GmbH + Co. KG (Karlsruhe, DE)
3,3',5,5'-Tetramethylbenzidin	Carl Roth GmbH + Co. KG (Karlsruhe, DE)
Acetic acid	Carl Roth GmbH + Co. KG (Karlsruhe, DE)
Acetone	Carl Roth GmbH + Co. KG (Karlsruhe, DE)
Ampicillin sodium salt	AppliChem GmbH (Darmstadt, DE)
Bacto Tryptone	BD Biosciences (San Jose, CA, USA)
Bacto Yeast extract	BD Biosciences (San Jose, CA, USA)
Bovine Serum Albumin	PAN-Biotech GmbH (Aidenbach, DE)
Citric acid	Carl Roth GmbH + Co. KG (Karlsruhe, DE)
D(+)-Glucose monohydrate	Carl Roth GmbH + Co. KG (Karlsruhe, DE)
Dimethyl sulfoxide (DMSO)	Carl Roth GmbH + Co. KG (Karlsruhe, DE)
Di-sodium-hydrogenphosphate-di-hydrat (Na_2HPO_4 x $2H_2O$)	Carl Roth GmbH + Co. KG (Karlsruhe, DE)
dNTP mix	
Ethanol absolute	VWR International GmbH (Hannover, DE)
Ethylenediaminetetraacetic acid (EDTA)	Carl Roth GmbH + Co. KG (Karlsruhe, DE)
Glycerol	Carl Roth GmbH + Co. KG (Karlsruhe, DE)
Hydrochloric acid, 37 % (HCl)	Carl Roth GmbH + Co. KG (Karlsruhe, DE)
Hydrogen Peroxide	Carl Roth GmbH + Co. KG (Karlsruhe, DE)
Isopropyl-β-D-thiogalactopyranosid	Carl Roth GmbH + Co. KG (Karlsruhe, DE)
Kanamycin sulfate	Carl Roth GmbH + Co. KG (Karlsruhe, DE)
LE-Agarose	Biozym Scientific GmbH (Karlsruhe, DE)
Monopotassium phosphate (KH_2PO_4)	Carl Roth GmbH + Co. KG (Karlsruhe, DE)
Polyethylene Glycol	Carl Roth GmbH + Co. KG (Karlsruhe, DE)
Potassium chloride (KCl)	Carl Roth GmbH + Co. KG (Karlsruhe, DE)
Potassium citrate	Carl Roth GmbH + Co. KG (Karlsruhe, DE)
Recovery Medium	Lucigen Corporation (Middleton, WI, USA)
Sodium Chloride (NaCl)	Carl Roth GmbH + Co. KG (Karlsruhe, DE)
Spam-Agar	Hellmuth Carroux GmbH & Co. KG (Hamburg, DE)
Sulfuric acid	Carl Roth GmbH + Co. KG (Karlsruhe, DE)
Tetracycline hydrochloride	AppliChem GmbH (Darmstadt, DE)
Tris(hydroxymethyl)aminomethane (TRIS)	Carl Roth GmbH + Co. KG (Karlsruhe, DE)
TRIzol reagent	Thermo Fisher Scientific (Dreieich, DE)
Tween20	Carl Roth GmbH + Co. KG (Karlsruhe, DE)

2.1.4 Bacterial strains, phage and cell lines

Bacterial strains, phage and mammalian cell lines used in this study are listed in Table 2-4.

Table 2-4: *Bacterial strains, phage and cell lines*

Strain	Genotype/description	Application	Source
E. coli SS320	F'[proAB lacIqZ ΔM15 Tn10 (TetR)] araD139 Δ(ara-leu)7696 galE15 galK16 Δ(lac)X74 rpsL (StrR) hsdR2 (rK– mK+) mcrA mcrB1	Cloning of antigen library	Lucigen Corporation (Middleton, WI, USA)
E. coli TG1	[F' traD36 proAB lacIqZ ΔM15] supE thi-1 Δ(lac-proAB) Δ(mcrB-hsdSM)5(rK - mK -)	Panning	Lucigen Corporation (Middleton, WI, USA)
E. coli XL1 blue MRF'	Δ(mcrA)183 Δ(mcrCB-hsdSMR-mrr)173 endA1 supE44 thi-1 recA1 gyrA96 relA1 lac [F′ proAB lacIqZΔM15 Tn10 (Tetr)]	Panning and cloning	Stratagene (La Jolla, CA, USA)
HCT116	Human colorectal carcinoma cell line, growth in adhesion	RNA extraction	
Hyperphage	M13K07DgIII	Phage production	Rondot et al. 2001[42]

2.1.5 Plasmids

Plasmids used in this study are listed in Table 2-5.

Table 2-5: *Plasmids*

Plasmid	Description	Reference
pHORF3	Phagemid, coding for peptide fusions with minor coat protein pIII of bacteriophage M13, library vector for ORFeome display libraries	Kügler et al. 2008[43]
pTSD3	Phagemid, coding for peptide fusions with minor coat protein pIII of bacteriophage M13, library vector for unidirectional ORFeome display libraries	
pCSE2.6-mIgG2a-Fc-XP	Plasmid, coding for peptide fusions with mouse IgG2a Fc part, used for production of scFv-Fc antibodies in mammalian cell culture	Miethe et al. 2015[44]

2.1.6 Kits

Kits used in this study are listed in Table 2-6.

Table 2-6: *Kits*

Kit	Application	Manufacturer
NEBNext® Ultra™ II Directional RNA Library Prep Kit for Illumina®	cDNA library generation	New England Biolabs (Frankfurt a. M., DE)
NEBNext® Poly(A) mRNA Magnetic Isolation Module	mRNA purification	New England Biolabs (Frankfurt a. M., DE)
NucleoSpin® Gel and PCR Clean-up	DNA purification	Macherey-Nagel GmbH & Co. KG (Düren, DE)
NucleoSpin® Plasmid Transfection-grade	Plasmid extraction	Macherey-Nagel GmbH & Co. KG (Düren, DE)
NucleoBond® Xtra Midi	Plasmid extraction	Macherey-Nagel GmbH & Co. KG (Düren, DE)
Direct-zol RNA Miniprep Plus	total RNA isolation	Zymo Research Europe GmbH (Freiburg, DE)
EasySep™ Human B Cell Isolation kit	B cell isolation	Stemcell Technologies (Vancouver, CAN)

2.1.7 Antibodies

Antibodies used in this study are listed in Table 2-7.

Table 2-7: *Antibodies*

Antibody	Description	Manufacturer
11973-MM05T-H	Anti-M13 Antibody (HRP)	Sino Biological Europe GmbH (Eschborn, DE)
GSM238-B7	anti-MTP64 (*M. tuberculosis*) antibody	AG Dübel (Braunschweig, DE)
MA1-21315-HRP	6x-His Tag Monoclonal Antibody (HIS.H8), HRP	Thermo Fisher Scientific Inc. (Waltham, USA)
SUW57-D11	anti-mesothelin antibody	AG Dübel (Braunschweig, DE)

2.1.8 Oligonucleotides

Oligonucleotides were used for the construction of plasmids, as adapters in cDNA library generation and as primers for DNA amplification via PCR. All oligonucleotides used in this study are listed in Table 2-8.

Table 2-8: *Oligonucleotides*

Name	ID	Sequence 5' → 3'	Application
BtgZI_mRNA_hp_fw	YP559	ACAGGGTTTGGGAGATAACA	amplification of library inserts
BtgZI_mRNA_hp_rev	YP560	GTGAAACCCAAAGTCTCTCC	amplification of library inserts
MHgIII_r	YP11	CTAAAGTTTTGTCGTCTTTCC	colony PCR on pTSD3
MHLAacZ-Pro_f	YP125	GGCTCGTATGTTGTGTGG	colony PCR on pTSD3
BtgZI_mRNA_hairpin	YP558	GCCATCGTGGCTTCATCGCGGAGAGACTTTGGGTTTCACUACAGGGTTTGGGAGATAACAGCGATGTCGTAGCCACGATGGC*T	ligation to cDNA inserts during library generation
pHORF_BsmBI_leader	YP432	CTAGCAGCTCAGCCGGCTATGGCGAGAGACGATTGATGACGCGTCTCGCATCATCACCATCATCATTAGG	construction of pTSD3
pHORF_BsmBI_reverse	YP433	GATCCCTAATGATGATGGTGATGATGCGAGACGCGTCATCAATCGTCTCTCGCCATAGCCGGCTGAGCTG	construction of pTSD3
Lac_Op-for		CGGATAACAATTTCACACAG	sequencing of pHORF3 and pTSD3 vectors at Microsynth Seqlab GmbH (Göttingen, DE)
*: phosphorothioate bond for increased stability			

2.1.9 Enzymes, buffers and markers

The enzymes, commercial buffers and markers used in this study are listed in Table 2-9.

Table 2-9: *Enzymes, buffers and markers*

Application	Enzyme / Buffer / Marker	Manufacturer
Buffers		
restriction of DNA	NEBuffer 3.1	New England Biolabs (Frankfurt a. M., DE)
colony PCR	5x Green GoTaq® Reaction Buffer	Promega GmbH (Mannheim, DE)
restriction of DNA	CutSmart Buffer	New England Biolabs (Frankfurt a. M., DE)
Endonucleases		
restriction of DNA	BamHI HF	New England Biolabs (Frankfurt a. M., DE)
digestion of pTSD3 vector	BsmbI	New England Biolabs (Frankfurt a. M., DE)
restriction of DNA	BtgZI	New England Biolabs (Frankfurt a. M., DE)
restriction of DNA	NheI HF	New England Biolabs (Frankfurt a. M., DE)
digestion of pHORF3 vector	PmeI	New England Biolabs (Frankfurt a. M., DE)
Ligases		
ligation of inserts into vector backbones	T4 DNA Ligase	Promega GmbH (Mannheim, DE)
Markers		
agarose gel electrophoresis	Gel loading Dye Purple 6x	New England Biolabs (Frankfurt a. M., DE)
agarose gel electrophoresis	GeneRuler 1 kB Plus DNA ladder	Thermo Fisher Scientific (Dreieich, DE)
agarose gel electrophoresis	HDGreen DNA stain	Intas Science Imaging Instruments GmbH (Göttingen, DE)
Polymerases		
colony PCR	GoTaq® DNA-Polymerase	Promega GmbH (Mannheim, DE)
Others		
dephosphorylation of restricted DNA vectors	Alkaline Phosphatase, Calf Intestinal (CIP)	New England Biolabs (Frankfurt a. M., DE)
tissue digestion	Collagenase G	Sekisui Diagnostics LLC (Burlington, MA, USA)
tissue digestion	Collagenase H	Sekisui Diagnostics LLC (Burlington, MA, USA)
tissue digestion	DNase I	Sigma-Aldrich (Taufkirchen, DE)
elution of phage	Trypsin	Sigma-Aldrich (Taufkirchen, DE)

2.1.10 Media, Supplements and Buffers

Media, solutions and supplements for mammalian cell culture and cultivation of *E. coli* are listed in Table 2-10, 2-11 and 2-12. Supplements were added under sterile conditions.

Table 2-10: *Media, solutions and supplements for mammalian cell culture*

Media / Solution	Application	Manufacturer
DMSO	freezing of HCT116 cells	Biochrom GmbH (Berlin, DE)
FBS superior (Fetal Bovine Serum)	supplement for cultivation of HCT116 cells	Biochrom GmbH (Berlin, DE)
McCoy's 5a Medium	cultivation of HCT116 cells	Sigma-Aldrich, Inc. (St. Louis, MO, USA)
Penicillin / Streptomycin	cultivation of HCT116 cells	Biochrom GmbH (Berlin, DE)
Trypsin/EDTA solution	detachment of cells	Biochrom GmbH (Berlin, DE)

Table 2-11: *Basic media for cultivation of bacteria*

Media	Ingredient	Composition
2YT media	Bacto Tryptone	1.6 % (w/v)
	Bacto Yeast extract	1.0 % (w/v)
	NaCl	0.5 % (w/v)
	Agar (for solid media)	1.5 % (w/v)
LB media	Bacto Tryptone	1.0 % (w/v)
	Bacto Yeast extract	0.5 % (w/v)
	NaCl	1.0 % (w/v)
	Agar (for solid media)	1.5 % (w/v)
SOB media	Bacto Tryptone	2.0 % (w/v)
	Bacto Yeast extract	0.5 % (w/v)
	NaCl	0.5 % (w/v)
	KCl	0.02 % (w/v)
SOC media	SOB medium	490 ml
	MgCl2 (2 M)	5 ml
	Glucose (2 M)	5 ml

Table 2-12: *Supplements for cultivation of bacteria*

Supplement	Stock concentration	Final Concentration
Ampicillin (A)	100 mg/ml	100 µg/ml
Tetracycline (T)	10 mg/ml	20 µg/ml
Kanamycin (K)	50 mg/ml	100 mM
Glucose (G)	2 M	100 mM
IPTG	1 M	50 µM

2.1.11 Buffers and solutions

Non-commercial buffers and solutions are listed below (Table 2-13). If not stated otherwise, pH was adjusted using 1 M HCl and 1 M NaOH.

Table 2-13: *Non-commercial buffers and solutions*

Buffer/Solution	Ingredients		Solvent
10x GA	Glucose, 2 M	49.5 % (v/v)	
	Ampicillin, 100 mg/ml	1 % (v/v)	
	2YT medium	49.5 % (v/v)	
Agarose gel	Agarose	1.5 % (w/v)	TAE buffer
	HDGreen	0.002 % (v/v)	
BSA-PBST	BSA	2 % (w/v)	PBST
FC-Buffer	FBS superior	2 % (v/v)	PBS
	EDTA	1 mM	
Glucose	Glucose	2 M	Milli-Q water
Glycerol	Glycerol	80 % (v/v)	Milli-Q water
Hydrochloric acid (HCl)	HCl	5 M	Milli-Q water
Magnesium Chloride	$MgCl_2$	2 M	Milli-Q water
Milli-Q-Tween	Tween20	0.05 % (v/v)	Milli-Q water
PBST	Tween20	0.05 % (v/v)	PBS
PEG/NaCl	Polyethylene glycol	20 %	Milli-Q water
	NaCl	2.5 M	
Phage dilution buffer (pH 7.5)	TRIS-HCl (1 M)	1 % (v/v)	Milli-Q water
	EDTA	2 mM	
	NaCl	20 mM	
Phosphate buffered saline (PBS)	NaCl	0.8 % (w/v)	Milli-Q water
	KCl	0.02 % (w/v)	
	Na_2HPO_4 x $2H_2O$	0.14 % (w/v)	
	KH_2PO_4	0.024 % (w/v)	
TAE buffer	TRIS	40 mM	Milli-Q water
	EDTA	2 mM	
	Acetic Acid	20 mM	
TMBA	Potassium citrate	30 mM	Milli-Q water
	Citric acid	50 mM	
TMBB	Acetone	10 % (v/v)	
	Ethanol	89.7 % (v/v)	
	Hydrogen peroxide	0.3 % (v/v)	
	Tetramethylbenzidine	1 mM	
TMB	TMBA	95 % (v/v)	
	TMBB	5 % (v/v)	
TRIS-HCl (pH 7.5)	TRIS	1 M	Milli-Q water
	HCl, 5 M	to pH 7.5	
Trypsin	Trypsin	10 µg/ml	PBS

2.1.12 Software

Software used in this study is listed in Table 2-14.

Table 2-14: *Software*

Software	Application	Reference
Citavi	Citation	Swiss Academic Software GmbH (Wädenswil, CHE)
Graph Pad Prism 7	Figures	GraphPad Software (San Diego, CA, USA)
Inkscape 0.91	Figures	https://inkscape.org/
Microsoft office	Text and data editing	Microsoft Corporation (Redmond, WA, USA)
NCBI BLASTn	Identification of cDNA insert sequences	https://blast.ncbi.nlm.nih.gov/Blast.cgi
Notepad++	Python script editing	https://notepad-plus-plus.org/
SWISS-MODEL	Protein structure homology modelling	Waterhouse et al. 2018)
Ugene	in silico cloning, sequence analysis, sequence alignment	UniPro, Novosibirsk, RU
Vortex	Data analysis, figures	CM Labs Simulations (Montreal, CAN)

2.2 Methods

2.2.1 Ethics

Experimental plans and patient information material were approved by the ethics committee of the TU Braunschweig. Patients were briefed on the scope and aims of this study by their physician prior to giving their consent. None of the conducted procedures and the collected data allow conclusions about the patient's identity.

2.2.2 Tissue handling

2.2.2.1 Samples

Fresh head and neck tumor dissections in 0.9 % saline solution were obtained from the Hals-Nasen-Ohren-Klinik of the Städtisches Klinikum Braunschweig. Tissues were processed in the lab no later than 2 h after dissection. Samples were collected without special regards to age and sex of the patients or stage of disease.

2.2.2.2 Creation of a single cell suspension

First, tissues were manually minced with small surgical scissors. Then, tissues were subjected to an enzymatic digest with 10 U Collagenase H, 3.4 U Collagenase G and 0.3 U DNase I for 1 h at 37 °C in FC buffer. Afterwards tissues pieces were pushed through a 40 µm cell strainer (Fisher Scientific GmbH, Schwerte, DE) with a syringe plunger and the strainer was flushed with 2 ml FC buffer.

2.2.2.3 B cell separation and storage

B cells were isolated from single cell suspensions (see section 2.2.2.2) using the EasySep™ Human B Cell Isolation kit (Stemcell Technologies, Vancouver, CAN), which works by CD19 negative selection, according to manufacturer's instructions. The non-selected fractions of cells, containing all CD19 negative cells, were also collected. CD19 B cells and pooled non-B cell fractions were spun down at 300 xg for 5 min, supernatants were discarded and 1 ml of TRIzol reagent was added per up to 10^7 cells. Samples were stored at -80 °C up to 12 months. B cells were used to generate recombinant antibody libraries in the course of a thesis closely related to this one (Melanie Philippi).

2.2.3 Molecular biological methods

2.2.3.1 DNA extraction

Plasmid DNA from *E. coli* cultures was isolated using the NucleoSpin® Plasmid or the NucleoBond® Xtra Midi kit (Macherey-Nagel GmbH & Co. KG, Düren, DE), according to manufacturer's instructions.

2.2.3.2 RNA extraction

Whole RNA from tissue and cell line samples dissolved in TRIzol reagent was extracted using the Direct-zol RNA Miniprep Plus kit (Zymo Research Europe GmbH, Freiburg, DE), according to manufacturer's instructions.

2.2.3.3 Agarose gel electrophoresis

DNA vectors, digested DNA vectors and fragments and PCR amplification products were analysed and purified on agarose gels. 1.5 % (w/v) agarose gels containing 1:50.000 (v/v) HDGreen DNA stain (Intas Science Imaging Instruments GmbH, Göttingen, DE) were run in TAE buffer at 130 V for 25-40 min. As a marker GeneRuler 1 kB Plus DNA Ladder Mix (Thermo Fisher Scientific, Dreieich, DE) was used. Gels were imaged on a ChemiDoc imaging system (Bio-Rad Laboratories, München, DE). For purification of fragments, gel pieces were cut out with a ceramic scalpel.

2.2.3.4 DNA purification

Digested DNA vectors, DNA fragments and PCR products were purified directly from the reaction mix or after separation on an agarose gel using the PCR clean-up kit (Macherey-Nagel GmbH & Co. KG, Düren, DE), according to manufacturer's instructions.

2.2.3.5 Colony PCR

Colony PCR was performed after library cloning to estimate the insert rate i.e. percentage of plasmids that carry an insert or during cloning to select positive clones for sequence confirmation. For the pTSD3 vector, primers YH11 and YH125 were used. Colony PCR reactions were setup as follows (Table 2-15).

Table 2-15: *Composition of colony PCR reaction*

Component	Final Concentration
5x Green GoTaq reaction buffer	1x
10 mM dNTP mix	200 µM
forward primer	500 nM
reverse primer	500 nM
GoTaq2 DNA Polymerase	0.25 U/µl
Milli-Q H_2O	Up to 10 µl

The thermal cycler was set to the following program (Table 2-16).

Table 2-16: *Thermal cycler program for colony PCR*

Phase	Temperature	Time	
Initial	95 °C	2 min	
Denaturation	95 °C	30 s	
Annealing	55 °C	20 s	30 cycles
Elongation	72 °C	30 s	
Final Elongation	72 °C	5 min	

2.2.3.6 DNA Sanger Sequencing

To confirm successful cloning and to identify antigen inserts selected during panning, DNA sequencing was performed. Bacterial cultures or plasmid DNA from the appropriate single bacterial clones were prepared and sent to Microsynth Seqlab GmbH in Göttingen for Sanger sequencing.

2.2.4 Cloning

2.2.4.1 Construction of pTSD3

The pTSD3 vector was constructed form pHORF3. pHORF3 vector restriction digest was composed as described in Table 2-17.

Table 2-17: *Composition of pHORF3 vector restriction digest reaction*

Reagent	amount
pHORF3 plasmid DNA	5 µg
CutSmart buffer	2 µl
NheI HF	1 µl
BamHI HF	1 µl
Milli-Q H_2O	Up to 20 µl

The reaction was carried out at 37 °C for 1 h. 1 µl of CIP was added, followed by brief vortexing and incubation at 37 °C for 1 h. The reaction was inactivated by incubation at 80 °C for 20 min. Plasmid digestion was checked and purified by agarose gel electrophoresis. To this end, the whole volume was mixed with the appropriate amount of 6x Purple Gel Loading Dye (New England Biolabs, Frankfurt a. M., DE). The band corresponding to the digested vector was cut from the gel under UV light and plasmid DNA was extracted.

Oligonucleotides YH432 and YH433 were mixed in a molar ratio of 1:1 and diluted in Milli-Q H2O for a final concentration of 100 nM. The sample was heated to 90 °C for 5 min and slowly cooled to room temperature to allow annealing of the complementary strands. Ligation of the oligonucleotide fragment into the pHORF3 backbone was carried out at

16 °C overnight followed by heat inactivation at 65 °C for 10 min. Ligation reaction was composed as described in Table 2-18.

Table 2-18: *Composition of the ligation reaction for pTSD3 construction*

Reagent	amount
annealed oligonucleotides (100 nM)	0.6 µl
T4 DNA ligase buffer (10x)	2 µl
digested pHORF3 vector backbone (BamHI HF & NheI HF)	50 ng
T4 DNA ligase	1 µl
Milli-Q H2O	Up to 20 µl

10 µl of ligation reaction were added to 25 µl of chemically competent *E. coli* XL1 blue MRF' bacteria and incubated on ice for 20 min. Heat shock was applied at 42 °C for 1 min. 165 µl of pre-warmed SOC medium were added and the culture was incubated at 37 °C and 650 rpm for 1 h. The complete volume was spread on 2YT-GA agar plates which were incubated at 37 °C overnight.

2.2.4.2 cDNA fragment generation for antigen libraries

Using 1 µg freshly isolated RNA (see section 2.2.3.2) as starting material, the NEBNext® Poly(A) mRNA Magnetic Isolation Module (New England Biolabs, Ipswich, MA, USA) was used in combination with the NEBNext® Ultra™ II Directional RNA Library Prep Kit for Illumina® (New England Biolabs, Ipswich, MA, USA) to generate strand-specific cDNA libraries, according to manufacturer's instructions. As suggested, SPRIselect (Beckman Coulter GmbH, Krefeld, DE) magnetic beads were used for all nucleotide purification steps. The following changes were made to the protocol:

1. For adapter ligation a newly designed oligonucleotide (YH558) was used to introduce BtgZI restriction sites on both sides of the insert.
2. To test the creation of libraries with different insert sizes (see section 3.2.1), RNA fragmentation time was adjusted to 7 min or 5 min instead of 15 min.
3. For PCR Enrichment of Adaptor ligated DNA, primers YH559 and YH560 were used. PCR was run for 15 cycles.

The resulting amplified cDNA fragments were either stored at -20 °C for a maximum of 24 h or immediately used for cloning into pTSD3 vector.

2.2.4.3 Cloning of tumor tissue cDNA fragments into pTSD3 vector

pTSD3 vector was digested in larger batches and used for several library cloning experiments. The restriction digest was composed as described in Table 2-19.

Table 2-19: *Composition of bulk pTSD3 vector restriction digest reaction*

Reagent	amount
pTSD3 plasmid DNA	25 µg
NEBuffer 3.1	10 µl
BsmBI	2 µl
Milli-Q H_2O	Up to 100 µl

The reaction was carried out at 55 °C for 2 h. 1 µl of CIP was added, followed by brief vortexing and incubation at 37 °C for 1 h. BsmBI was inactivated by incubation at 80 °C for 20 min. Plasmid digestion was checked and purified by agarose gel electrophoresis. The whole volume was mixed with the appropriate amount of 6x Purple Gel Loading Dye (New England Biolabs, Frankfurt a. M., DE) and separated on a 1.5 % agarose gel running at 130 V in NE buffer for 35 min. The band corresponding to the digested vector was cut from the gel with a scalpel under UV light and plasmid DNA was extracted.

Library cDNA fragments were digested at 60 °C for 2 h with the reaction composition described in Table 2-20.

Table 2-20: *Composition of cDNA fragment restriction digest reaction*

Reagent	amount
amplified cDNA fragments	20 µl
CutSmart buffer	5 µl
BtgZI	1 µl
Milli-Q H2O	to 50 µl

Digested fragments were purified using SPRIselect beads (Beckman Coulter GmbH, Krefeld, DE) according to manufacturer's instructions by applying 90 µl of beads. Ligation of cDNA fragments into pTSD3 vector was carried out at 16 °C for 5 h followed by heat inactivation at 65 °C for 10 min. The reaction was composed as described in Table 2-21.

Table 2-21: *Composition of cDNA fragment ligation reaction*

Reagent	amount
purified cDNA fragments	20 µl
T4 DNA ligase buffer (10x)	10 µl
digested pTSD3 vector backbone (BsmBI)	100 ng
T4 DNA ligase	1 µl
Milli-Q H2O	to 100 µl

Prior to electroporation, ligation reactions were subjected to desalting using Amicon Ultra-0.5 Centrifugal Filter Unit with a 30 kDa cut-off (Merck KGaA, Darmstadt, DE) to remove unwanted salts and facilitate successful transformation. 100 µl of ligation reaction were added to the column. The volume was brought to 400 µl by adding Milli-Q H2O and the column was spun at 16.000 xg for 10 min. This was repeated two more times for a total of three centrifugations. Concentrated, desalted, ligated DNA was eluted from the column by placing it upside down in a new collection tube and spinning it at 2.000 xg for 10 seconds.

2.2.4.4 Transformation of electrocompetent *E. coli*

SS320 electrocompetent *E. coli* (Lucigen Corporation, Middleton, WI, USA) were used for electroporation. 25 µl of cells were mixed with 10 µl of ligation reaction. The mixture was placed into pre-chilled 0.1 cm GenePulser cuvettes (Bio-Rad Laboratories, München, DE) and pulsed with 1.8 kV on a BioPulser system (Bio-Rad Laboratories, München, DE). 1 ml Recovery medium (Lucigen Corporation, Middleton, WI, USA) pre-warmed to 37 °C were added immediately and the cells were transferred to a new tube. Bacteria were incubated at 650 rpm and 37 °C for 1 h in a thermal mixer (Thermo Fisher Scientific, Dreieich, DE).

10 µl of electroporated cells were taken for estimation of transformation efficiency. The remaining bacterial culture was spread on 2YT-GAT pizza plates and grown overnight at 37 °C. 35 ml of 2YT medium were added and plates were rocked on a sample rocker for 30 min. Bacteria were scraped off with an L-spatula. The liquid culture was collected and used to create 12 glycerol stocks. From the remaining bacteria, plasmid DNA was isolated. Insert rates were estimated by performing colony PCR on 24 clones from the 2YT-GA plates used for estimation of transformation efficiency.

2.2.4.5 Estimation of transformation efficiency

In 2YT medium, electroporated cells were diluted and spread on 2YT-GA agar plates for final dilutions of 10^{-4} and 10^{-6}. Colonies were grown overnight at 37 °C. Colonies were counted and number of transformants calculated according to the dilution factor.

2.2.5 Bacterial cultivation techniques

2.2.5.1 Sterilization

Prior to use, all media and supplements were sterilized in an autoclave at 121 °C and 1 bar for 20 min. Temperature sensitive liquids were passed through a 0.2 µm filter.

2.2.5.2 Bacterial cultures

Bacteria cultures on agar plates were incubated at 37 °C in a standing incubator. Liquid cultures were incubated 37 °C in flasks or microtiter plates in shaking incubators at 250 rpm and 800 rpm respectively.

2.2.5.3 Storage of bacterial cultures (glycerol stocks)

For long-term storage, bacterial cultures were supplemented with 80 % glycerol stock solution to achieve a final concentration of 20 %, mixed well and stored at -80 °C.

2.2.5.4 XL1 and TG1 cultures for infection

30 ml 2YT medium were inoculated from 1 ml TG1 glycerol stocks with an OD_{600} of about 0.7. 30 ml 2YT-T medium were inoculated from 1 ml XL1 Blue MRF' glycerol stocks with an OD_{600} of about 0.7. Cultures were grown at 37 °C and 250 rpm until an OD_{600} of 0.4-0.5. If not otherwise specified, infection was carried out at 37 °C for 30 min with and then 30 min without shaking at 250 rpm or 300 rpm in a ZWYC-290A incubator (LABWIT Scientific Pty. Ltd, Melbourne, AUS) (for flasks and 96-well MTPs, respectively) or 500 rpm or 800 rpm in a Vortemp56 incubator (Labnet International, Inc, Edison, NJ, USA) (for 24-deepwell plates and 96-well MTPs, respectively).

2.2.5.5 Production of monoclonal peptide-pIII fusion proteins for screening ELISA

Randomly chosen bacterial clones that were selected during panning were used to inoculate 150 µl of 2YT-AT-IPTG medium in 96-well microtiter plates. Cultures were grown at 300 rpm and 30 °C overnight.

2.2.6 Phage handling

2.2.6.1 Hyperphage packaging for ORF enrichment

To present the peptides that are contained in the libraries on the surface of phage particles, the libraries were packaged using Hyperphage[42]. Library packaging was performed as described by Fühner et al.[35] with minor changes to the protocol. Initial inoculation was done with 1 ml library glycerol stock preparation in 400 ml 2YT-GAT medium in 1 l baffled flasks (step 4. under section 3.3). The following phage particle production was carried out in 600 ml 2YT-AK medium in 2 l non-baffled flasks at 25 °C (step 7. under section 3.3). Phage particle pellets were taken up in phage dilution buffer instead of PBS.

2.2.6.2 Titration

Dilutions for titration were done using PBS. Titers of Hyperphage packaged libraries and amplified and eluted phage particle solutions were determined as described before[45].

2.2.6.3 Panning

2.2.6.3.1 Surface panning in MTP

Antigen panning in MTP format was performed as described by Fühner et al.[35] with the following changes to the protocol. Instead of Panning Block solution, 2 % BSA PBST were used. Preincubation with a negative control antibody was omitted. XL1 Blue MRF' and TG1 *E. coli* cultures were inoculated from glycerol stocks on the day of use, not from overnight cultures. A second and third panning round were performed. For amplification of phage particles, the whole elution volume from the previous round was used for infection and Hyperphage was used for polyvalent display (step 3 and 4 under section 3.7). The third panning round was stopped after elution of phage particles.

2.2.6.3.2 Panning in solution

All 1.5 ml reaction tubes used for panning were blocked with 2 % BSA-PBST for at least 1 h prior to use. Blocking solution was removed by pipetting. 300 µl 2 % BSA-PBST were placed in a tube. An equivalent of at least 50x the estimated diversity of the library of peptide presenting Hyperphage particles (in most cases 5 x 10^8) was added, followed by addition of 200 ng of the desired antibody for selection. For patient serum, 50 µl of a 1:1000 dilution in PBS were added. The panning reaction was incubated while rotating at 15 rpm on a Multi Bio RS-24 rotator (SIA Biosan, Riga, LV) for 2 h. Antibodies were captured from the panning solution using SureBeads™ Protein A Magnetic Beads (Bio-Rad Laboratories, München, DE). 15 µl of beads for each panning reaction were prepared according to manufacturer's instructions and added to the mixture. After 10 min incubation while rotating at 15 rpm, tubes were placed on a DynaMag™-2 magnetic rack (Thermo Fisher Scientific, Dreieich, DE) and the supernatant was discarded. Beads were washed five times with 800 µl PBST with 3 s vortexing and 30 s magnetization in between. After the second and third panning round, washing steps were increased to 10 and 15 times, respectively. Finally, beads were resuspended in 200 µl 10 µg/ml Trypsin solution and phage particles were eluted at 37 °C for 30 min. The tube was placed on the magnetic rack and the eluate was removed.

Amplification of phage particles with Hyperphage was carried out as described before[35] using the whole 200 µl of eluted phage particles.

2.2.6.3.3 Panning in 96-well plate

Conducting panning in a 96-well plate follows the same principles as described before (see section 2.2.6.3.2) with some adaptations due to the format. All used wells of a V-bottom 96-deepwell plate (Ritter GmbH, Schwabmünchen, DE) were blocked with 700 µl 2 % BSA PBST. Total volume of the panning mixture was reduced to 500 µl, using only 250 µl 2 % BSA PBST. Panning and bead capturing process were incubated in a Vortemp56 incubator Vortemp56 incubator (Labnet International, Inc, Edison, NJ, USA) at 25 °C and 650 rpm. Washing steps were performed using the mixing option of a 96-channel electric Viaflo96 pipette (INTEGRA Biosciences GmbH, Biebertal) with 300 µl PBST and a speed setting of 8 out of 10 for 3 cycles. A 96S Super Magnet Plate (Alpaqua Engineering LLC, Beverly, MA, USA) was used for magnetizations. The elution reaction was also incubated in a Vortemp incubator at 37 °C and 650 rpm.

2.2.6.4 Production of monoclonal phage particles for screening ELISA

Randomly chosen bacterial clones that were selected during panning were used to inoculate 150 µl of 2YT-GAT medium in 96-well microtiter plates. Cultures were grown at 300 rpm and 37 °C for 2 h. 5 x 10^8 Hyperphage particles were added and infection was carried out at 37 °C for 30 min without and then 30 min with shaking. Plates were then spun at 3220 xg for 10 min and supernatants were discarded. 200 µl of 2YT-AK medium were added and plates were incubated at 30 °C and 300 rpm overnight.

2.2.7 Immunological assays

2.2.7.1 Screening ELISA with monoclonal phage

Sample antibody mixture, either 20 ng oligoclonal scFv-Fc or 30 µl of patient serum diluted 1:4000 in PBS, were coated in 30 µl of PBS at 4 °C overnight. Antibody solutions were knocked out of the wells and wells were then blocked with 1 % BSA PBST for 1 h. 30 µl of phage production supernatant were added and incubated for 1 h. 30 µl anti-M13 antibody coupled to HRP (final concentration: 55 ng/µl) in PBS were added and incubated for 1 h. Between incubation steps wells were washed three times with Milli-Q PBST. Binding was visualised by addition of 30 µl 3,3′,5,5′-tetramethylbenzidine substrate and the reaction was stopped by adding 30 µl 1 N sulfuric acid. Using a Sunrise microtiter plate reader (Tecan Group AG, Männedorf, CHE), absorbance at 450 nm and scattered light at 620 nm were measured. For signal evaluation the scattered light intensity was subtracted from the absorbance value.

2.2.7.2 Screening ELISA with peptide-pIII fusions

Peptide-pIII fusion ELISA was, in principle, performed as described for phage particle ELISA. Phage production supernatants were replaced by supernatants of peptide-pIII fusion productions. For detection, 30 µl anti-His-tag antibody coupled to HRP (final concentration: 62.5 ng/µl) in PBS was used.

2.2.8 Next generation sequencing

2.2.8.1 sequencing library preparation

30 ml XL1 Blue MRF' *E. coli* cultures were infected with phage packaged unselected library with an 50x excess of library diversity. 4 ml XL1 Blue MRF' cultures were infected with phage particles that were amplified or eluted during panning, using 1 x 10^8 amplified phage particles or 100 – 200 µl of the elution solution. Bacteria were pelleted by centrifugation at 3220 xg, supernatants were discarded, and pellets were taken up in the same volume 2YT-GA medium. Cultures were grown at 37 °C and 250 rpm overnight. Plasmid DNA was isolated and sent to the EMD Serono Research & Development Institute for NGS analysis.

Sequencing libraries were kindly prepared by Thomas Clarke using either the MiSeq Reagent Kit v3 (150 cycles) or the NextSeq 500/550 Mid Output Kit v2.5 (150 cycles) from Illumina according to manufacturer's instructions. Sequences were read using an Illumina MiSeq or NextSeq500 system running for 75 or 150 cycles, respectively.

2.2.9 Computational methods

2.2.9.1 NGS sequence analysis

After base calling and quality scoring sequencing reads were passed through general quality control eliminating low quality reads from the data using the provided software on the Illumina BaseSpace platform. From sequences generated thereby, primer and pTSD3 vector sequences were trimmed away. Using the pseudoaligner kallisto, reads were aligned to genes and counted. For the estimation of unique sequences within the libraries, a custom python script was created. The script used .fastq files with sequences that passed quality control as input and trimmed away primer and pTSD3 vector sequences. A sequence was accepted as unique when it differed in length or sequence by at least 1, compared to all other sequences. Only forward reads were used for the analysis of unique sequences.

2.2.9.2 Sequence alignments

Sequence alignments were performed using the ClustalW algorithm embedded in Ugene (UniPro, Novosibirsk, RU) with gap opening penalty set to 70 and gap extension penalty set to 10.

2.2.9.3 BLASTn analysis

The identity of insert sequences selected during panning and screening is analysed by an NCBI BLASTn search against the human genomic and transcript reference sequences. Those sequences that have 100 % query cover and share at least 98 % identity with the reference are considered matches.

2.2.10 Cell biological methods

All work regarding mammalian cell culture was kindly performed by Melanie Philippi and Marie Kastull.

2.2.10.1 Cultivation of mammalian cells

HCT116 cells were cultivated according to ATCC guidelines, as stated on their website and used as a source of total RNA for antigen library generation.

2.2.10.2 Harvesting of HCT116 cells for RNA extraction

Supernatants were discarded. Cells were detached from the dish surface by applying 2 ml Trypsin/EDTA solution and incubation at 37 °C for 5-10 min. Cells were taken up in 10 ml McCoy's 5a Medium and counted in a Neubauer chamber (BRAND GMBH + CO KG, Wertheim, DE) under a IX70 light microscope (Olympus Europa SE & CO. KG, Hamburg, DE). 10^7 cells were pelleted by centrifugation at 300 xg for 5 min, supernatants were discarded, and the pellet was dissolved in 1 ml TRIzol reagent.

3 Results

The patient samples that were procured for this study received anonymous descriptors consisting of the word “Yuhan” (for Yumab – Head and Neck Cancer) and a number, for example Yuhan007. Throughout this work patients or patient samples will only be referred to via these descriptors.

The following descriptions of experiments and results sometimes refer to oligoclonal recombinant antibody mixtures (kindly provided by Melanie Philippi). These were not individually composed following a rationale. They were created using scFv antibody libraries derived from tumor infiltrating B cells (TIBC) that were isolated from patient tissue. The scFv fragments were transferred into the scFv-Fc expression vector pCSE2.6-mIgG2a-Fc-XP[44] via mass cloning and plasmid DNA preparations thereof were used to transfect HEK293-6E cells for production. Therefore, the exact content and stoichiometry of these mixtures are unknown. However, they represent the local antibody repertoire from within and around the tumor tissue.

3.1 Construction of the pTSD3 ORFeome display system

The previously described ORFeome display methods relied on DNA fragmentation by sonication and blunt end cloning into the phagemid vector pHORF3[43]. In first experiments using cDNA from cultured mammalian cells, sonication and blunt end cloning of cDNA fragments were not efficient and resulted in insert rates not exceeding 15 %. To improve cloning efficiency, a sticky end and unidirectional, coding stand-specific cloning procedure was tested. This strategy was also expected to double the number of inserts that are ligated with correct orientation to the gIII fragment. For this purpose, the new vector pTSD3 was constructed from the pHORF3 plasmid (Figure 3-1) by replacing the PmeI cloning site of pHORF3 with a stuffer fragment containing two asymmetric BsmBI type IIS restriction sites. This approach allows insertion of random cDNA fragments that have been ligated to the YH558 adapter. This adapter allows digestion with the restriction enzyme BtgzI to generate asymmetric, complementary overhangs for unidirectional cloning downstream of the 5’ PelB signal peptide and upstream of the 3’ His-tag and gIII sequence. By employing this new sticky-end cloning approach, insert rates of up to 95.8 % were achieved.

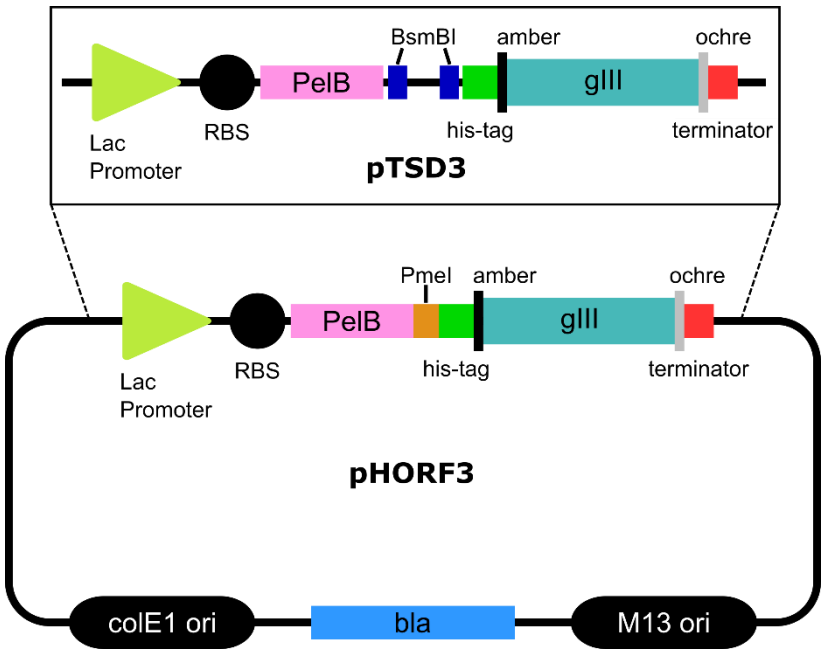


Figure 3-1 – pHORF3 and pTSD3: *Schematic vector map of pHORF3 and pTSD3 showing the changed cloning site. RBS: ribosome binding site; PelB: leader peptide; amber: amber stop codon (TAG); gIII: gene coding for phage pIII protein; ochre: ochre stop codon (*TAA*). (adapted from Kügler et al.[43])*

3.2 Antigen library analysis

3.2.1 Insert size for antigen display was limited to about 350 bp

Selection of peptides or protein fragments from antigen libraries against their corresponding antibodies (or antisera) greatly depends on their folding. While some antibodies bind short linear peptides, a significant portion requires the formation of a secondary or tertiary structure for correct antigen recognition[46]. Such antigens may only be identified if these conformations can be represented on the phage surface. The display of mammalian antigens on phage requires their secretory expression as pIII fusion proteins in the heterologous *E. coli* system. Bacteria lack the complex protein folding machinery of mammals and express larger proteins with low efficiency. However, many proteins are composed of several different protein domains which can form a correct tertiary structure despite representing only a fraction of the complete polypeptide chain of the protein. Such protein domains range in length from about 30 aa to 500 aa with the majority of domains having a size of less than 200 aa[47], which corresponds to 600 bases of coding mRNA sequence. Chains of less than 100 aa (300 coding bases) are mostly unfavourable for folding due to higher free energies of the unfolded states[48]. In order to increase the chance of full protein domains being present in the libraries, three HCT116 cell line libraries with intended average fragment sizes of 500, 350 and 200 bp were constructed (termed S1, S2 and S4) by adjusting the RNA fragmentation time in the library generation procedure. From the three resulting libraries, single clones were selected before and after packaging

with Hyperphage, sent for Sanger sequencing and analysed for their insert length. Reducing RNA fragmentation by lowering incubation time from the standard 15 min approach (S4) that aims for an average insert size of 200 bp to 7 min (S2) did not yield a noticeably different insert size distribution (Figure 3-2). Decreasing fragmentation time to 5 min yielded a library with a much broader insert size distribution, with 41.7 % of inserts exhibiting a length greater than 350 bp in library S1 compared to 13.7 % in library S4. Single inserts even reached lengths of around 850 bp. However, after Hyperphage packaging only 3.17 % of inserts with a length greater than 350 bp are present, showing that the packaging process in *E. coli* selects against larger pIII-fusion peptides. For this reason, all following libraries were created using the standard 15 min fragmentation protocol in order to achieve maximum diversity in the lower size cDNA fragments and to reduce the risk of losing antigen fragments that would drop out during Hyperphage packaging.

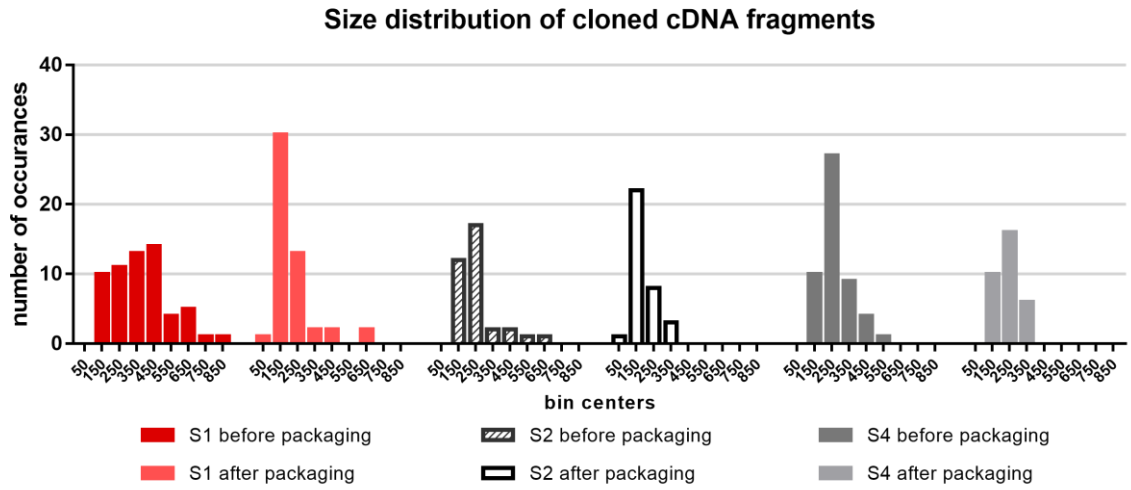


Figure 3-2 – Insert size distribution*: Histogram of cDNA insert size distributions for HCT116 libraries S1, S2 and S4 before and after packaging with Hyperphage. Libraries were generated with different RNA fragmentation times of 5 min (S1), 7 min (S2) and 15 min (S4)*

3.2.2 In frame ORF inserts are highly enriched after Hyperphage packaging

Depending on their length, randomly fragmented and unidirectionally cloned cDNA inserts can be present in one of three possible reading frames. In pTSD3, inserts that have a length divisible by three are in frame with the gIII sequence. The number of in frame constructs should greatly improve through packaging with Hyperphage, because only inserts without stop codon and without a reading frame shift are assumed to gain infectivity. To check ORF enrichment, single clones from the HCT116 S1, S2 and S4 libraries (see section 2.2.1) were analysed by Sanger sequencing before and for S1 and S4 also after packaging. Before packaging, 32.2 % of inserts are in frame, which is in accordance with the expected theoretical value of 33.3 % (Figure 3-3). After Hyperphage packaging the amount of in frame sequences increases to 81.7 %, demonstrating effective ORF enrichment. While roughly 11 % percent of sequences should contain ORFs that encode peptides aligning to human protein sequences prior to packaging, this is only the case for 2 % (3 out of 146 sequences). All other in frame sequences contained stop codons.

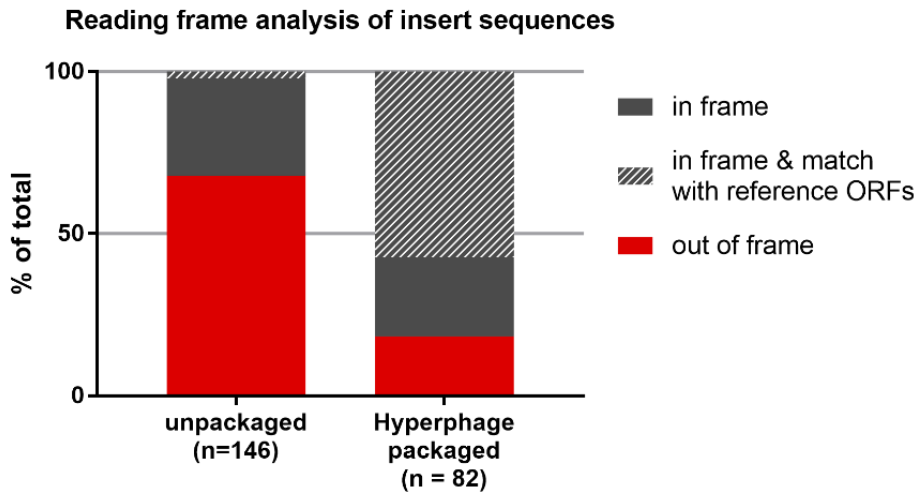


Figure 3-3 – ORF enrichment: *A total of 228 individual randomly picked clones from different HCT116 cell line pTSD3 libraries were sequenced before and after packaging with Hyperphage. Percentage of insert sequences that are in frame with the gIII sequence and percentage of sequences that match ORFs of proteins from the NCBI reference protein database was calculated.*

3.2.3 Next generation sequencing analysis of ORFeome libraries

Antigen display libraries were created from head and neck cancer tissue as well as from cancer cell lines. To get a deeper insight into the contents of these libraries, the Yuhan007, Yuhan008 and Yuhan011 patient libraries and the HCT116 S4 cell line library were analysed using next generation sequencing (NGS). To this end, plasmid DNA was isolated from *E. coli* XL1 blue MRF' bacteria after infection with phage packaged antigen libraries and used for NGS sequence analysis. All libraries were analysed in an Illumina MiSeq sequencing run, Yuhan007 and Yuhan008 were also part of a larger second sequencing run using the Illumina NextSeq platform.

The number of genes that are represented in the libraries with at least low coverage ranges from 3,541 to 7,169 (Table 3-1). Judging by the raw reads from the NextSeq run, the highest sequence diversity was reached at 3.7×10^6 unique sequences in 17,041,751 reads, after excluding empty vectors and reads containing incorrect vector backbone sequences. Sequences were termed unique when they differed at least by one base in length or sequence.

Aligning sequences that mapped to a single gene to the reference genome in the UCSC genome browser shows, that only exon sequences were present in these libraries (Figure 3-4 A and B).

The generated antigen libraries roughly covered up to one third of the human transcriptome. The exclusive presence of exon sequences demonstrates that the transcriptome library design was applicable for phage display libraries made from human RNA.

Table 3-1: *Gene diversity of cDNA libraries; number of genes that are represented in the antigen libraries with a read depth of more than 10, 100 or 1000 reads as determined by a MiSeq NGS run*

	Yuhan007	Yuhan008	Yuhan011	HCT116 S4
>10 reads	7169	3541	5760	6226
>100 reads	1381	408	717	822
>1000 reads	85	23	44	57

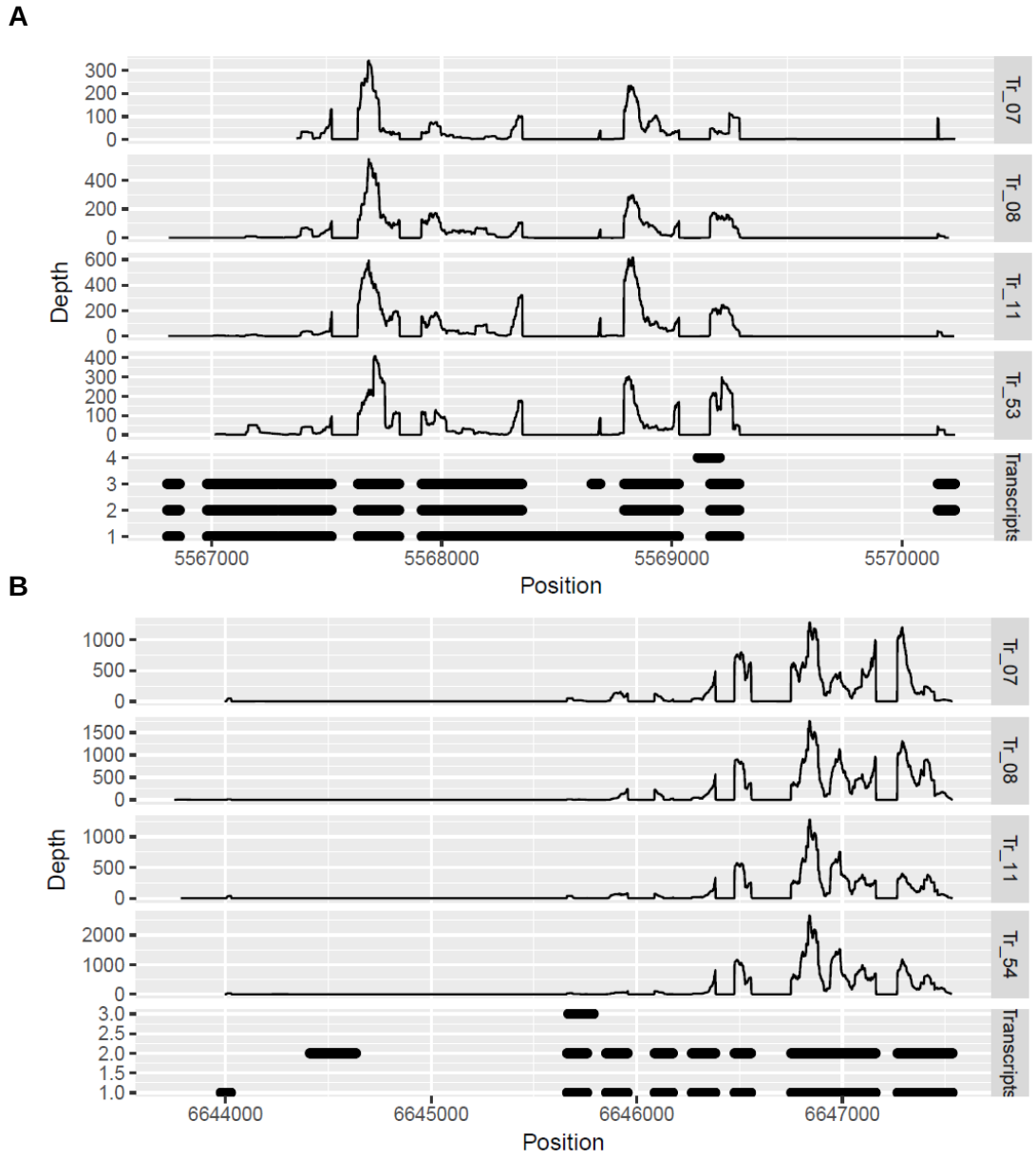


Figure 3-4 – Exon coverage: *Per base read depth of Beta-actin (ACTB) (A) and Glyceraldehyde-3-Phosphate Dehydrogenase (GAPDH) (B) sequences represented in four antigen libraries aligned to gene sequences listed in the UCSC genome browser. Tr_07: Yuhan007; Tr_08: Yuhan008; Tr_11: Yuhan011; Tr_53: HCT116 S4*

3.3 Technology validation

3.3.1 Binding peptides can be selected from antigen libraries using scFv-Fc antibodies

The aim of this work was to identify new polypeptide antigens that are bound by antibodies derived from the same patient tumor sample. In order to establish the procedures for the identification of antigenic polypeptides from the libraries using antibodies, phage display panning of the HCT116 S4 library (see section 2.2.1) was performed against the monoclonal human anti-mesothelin (MSLN) antibody SuW57-D11. From the amplified phage particles after the second panning round, 92 clones were analysed in form of secreted antigen-pIII fusion proteins by ELISA. Five out of 6 clones with signal to noise ratios greater than 4 and 5 compared to isotype and BSA control, respectively, were sent for Sanger sequencing (Figure 3-5). All five antigen clones contained MSLN sequences according to NCBI BLASTn, demonstrating that specifically bound protein fragments or peptides can be isolated from the antigen libraries via antigen phage display. The selected inserts were 216 or 219 bp long and identical in sequence except for three additional bases of the longer insert.

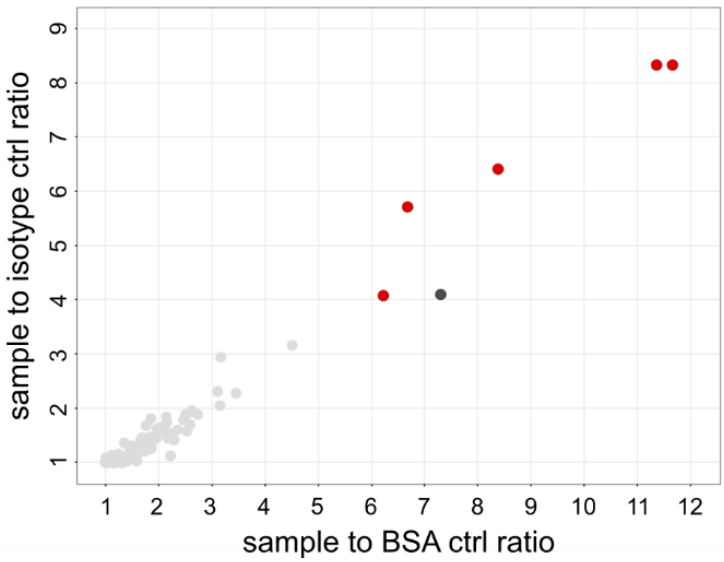


Figure 3-5 – Peptide screening ELISA for MSLN panning in MTP format: *Peptides selected by MTP panning were tested for their binding to anti-mesothelin antibody SuW57-D11 scFv-mouse-Fc in ELISA as antigen-pIII fusion proteins. Signal to noise ratios to the isotype control antibody GSM238-B7 scFv-mouse-Fc and to BSA are plotted. light grey dots: samples that did not pass the selection criteria; dark grey dots: samples that passed the selection criteria; red dots: samples that passed the selection criteria and identified as MSLN sequences after Sanger sequencing*

3.3.2 Panning in solution yielded more hit candidates than panning in MTP format

Phage display panning is often performed via immobilization of antigen or antibody on a surface and addition of a phage library of possible interaction partners[43,49]. However, adsorption to hydrophobic surfaces can have an impact on protein structure and binding to a surface reduces the accessible interaction area. Another possibility is to first allow both interaction partners to interact in solution and then capture them to a surface later, e.g. by using magnetic beads. To assess the significance of this factor for ORFeome display, panning on immobilized antibodies in MTP format was compared to panning in solution.

Two antigen phage display pannings were performed for each of the three HCT antigen libraries described in section 2.2.1. First, antibody SuW57-D11 in the scFv-mouse-Fc format was immobilized on the surface of an MTP and second, the same antibody was incubated with the antigen library in solution and later captured using protein A magnetic beads. After two rounds of panning, 92 randomly picked clones of each panning strategy and each library were screened by ELISA using the secreted antigen-pIII fusion proteins. Cut-off values for signal to isotype control and signal to BSA control ratios were set to 5 and 10, respectively, for both experiments. Panning in MTP format yielded only one positive clone whereas panning in solution yielded 41 positive clones from all three selections (Figure 3-6). Of these clones, 21 were sent for Sanger sequencing and all identified as MSLN mRNA sequences in BLASTn analysis. This comparison shows that panning in solution was more efficient for the selection of interaction partners in the current setting.

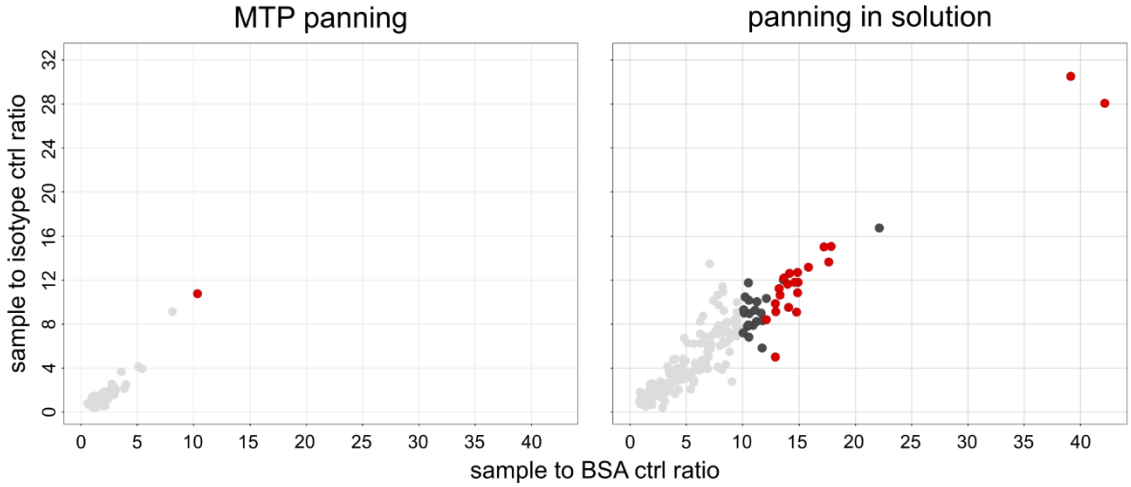


Figure 3-6 – Peptide screening ELISA for MSLN panning in two formats*: Peptides selected by MTP panning or panning in solution were tested for their specific binding to anti-mesothelin antibody SuW57-D11 scFv-mouse-Fc as secreted antigen-pIII fusion proteins in ELISA. Signal to noise ratios to the isotype control antibody GSM238-B7 scFv-mouse-Fc and to BSA are plotted. light grey dots: samples that did not pass the selection criteria; dark grey dots: samples that passed the selection criteria; red dots: samples that passed the selection criteria and identified as MSLN sequences after Sanger sequencing*

3.3.3 Phage ELISA versus antigen ELISA

Individual clones identified by panning can be screened for antibody binding by ELISA, either in form of secreted antigen-pIII fusion proteins or as antigen presenting phage particles. To check which approach performs better, the three times 92 clones picked from the panning in solution experiment described in 3.2 were subjected to antigen ELISA and phage ELISA in parallel. Cultivations for soluble expression of antigen-pIII fusion protein or Hyperphage packaging were inoculated from the same starting cultures. Non-purified cultivation supernatants were used for ELISA.

Compared to antigen ELISA, phage ELISA resulted in higher overall signal and improved resolution of data points, facilitating more confident hit selection. Applying the same cut-off values as before (see section 3.2), phage ELISA identified 77 positive clones (Figure 3-7). Of these, 51 were sent for Sanger sequencing and 47 identified as mesothelin mRNA sequences in BLASTn analysis, compared to 21 positive sequence identifications from 41 positive clones after antigen ELISA. Interestingly, there was only a small overlap of clones that are identified as MSLN mRNA sequences by both methods with 7 hits in 276 screened and 65 sequenced clones (Figure 3-8). A prevalence of different average insert sizes or of specific parts of the MSLN sequence was not observed for any of the two methods. Taken together these data suggest that phage ELISA is the favourable screening method for identifying hit candidates from antigen libraries using antibodies as capturing partner.

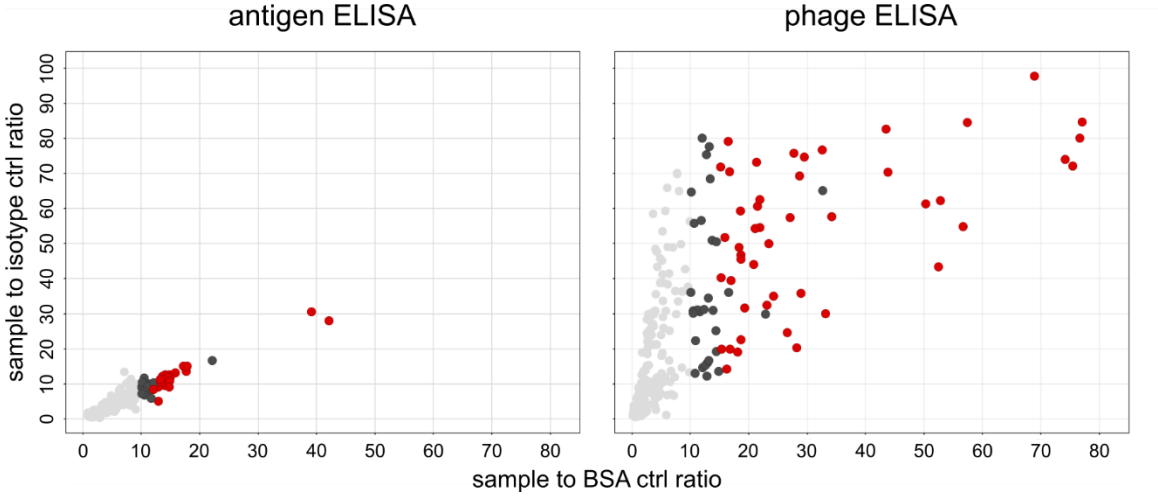


Figure 3-7 – Antigen versus phage screening ELISA: *Peptides selected by panning in solution were tested as antigen-pIII fusion proteins or Hyperphage packaged single clones for their specific binding to anti-mesothelin antibody SuW57-D11 scFv-mouse-Fc in ELISA. Signal to noise ratios to the isotype control antibody GSM238-B7 scFv-mouse-Fc and to BSA are plotted. Light grey dots: samples that did not pass the selection criteria; dark grey dots: samples that passed the selection criteria; red dots: samples that passed the selection criteria and identified as MSLN sequences after Sanger sequencing*

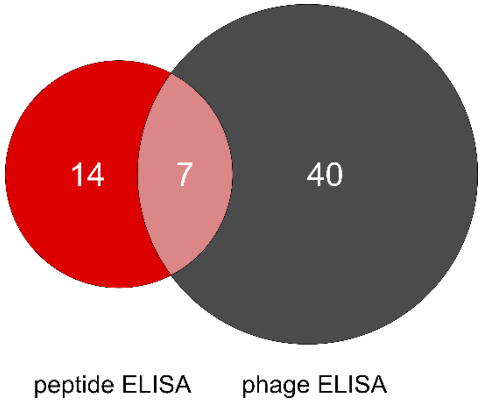


Figure 3-8 – Confirmed MSLN sequences: *Venn diagram showing the shares of confirmed MSLN sequences identified by antigen ELISA (red), phage ELISA (dark grey) and both methods (light red). 65 out of 276 screened clones were sequenced, 2 were not identified as MSLN sequences and 2 were disregarded due to low sequencing quality*

3.3.4 Panning in 96-well format

For reliable identification of targets that have a general relevance for patients in a given disease context, large sample numbers are needed. Performing panning of patient specific antigen libraries against the according patient antibody repertoires for dozens of patients is very time consuming. Even more so, if every patient library has to be cross selected with antibodies from every other patient individually. Therefore, it was tested if antigen panning in solution could also be performed in a 96 well-plate format to decrease overall handling time and make the method high throughput applicable without the need for equipment other than a 96-channel pipette and a magnetic plate.

A pool of antigen libraries was selected against recombinant oligoclonal antibody mixtures derived from TIBC from Yuhan011 and Yuhan012 (kindly provided by Melanie Philippi), against serum from Yuhan008 or against the anti-MSLN antibody SuW57-D11 (for control). The HCT116 S4 cell line antigen library was selected against SuW57-D11 as an additional control. A total of 192 clones from each selection were tested in phage ELISA. Clones were regarded as positive, when their signal to noise ratios for isotype and BSA control were greater than 5 and 10, respectively.

Selection of antigen fragments from Yuhan011, yielded 14 positive clones. Twelve of those were sequenced but did not show any enrichment of sequences that could be aligned to a single mRNA reference (Figure 3-9 A). This lack of enrichment made it impossible to choose a likely target candidate, since the enrichment should be an indicator for specific binding. For the Yuhan012 library, eight clones were positive in ELISA and 6 of those were sequenced. All sequences mapped to different mRNAs using BLASTn and were not showing enrichment of a certain gene. However, one of the sequences selected mapped to MMP9 which has also been enriched in other Yuhan012 selections (Figure 3-9 B). Out

of the 49 ELISA-positive clones from the Yuhan008 antigen library selection a total of 12 were sequenced. Despite very strong ELISA signals and high signal to noise ratios, no enrichment of sequences corresponding to a single gene could be seen (Figure 3-9 C). For the samples selected against anti-MSLN antibody SuW57-D11 all ELISA-positive and DNA sequenced clones (n = 5 and n = 6) could be mapped to MSLN mRNA (Figure 3-9 D and E).

Taken together, these data show that selection of binding peptides from patient derived cDNA fragment antigen libraries against monoclonal scFv-Fc antibodies by panning in solution is possible in a multi-sample paralleled approach. However, success of the selection process seems to greatly depend on the combination of starting library composition (i.e. single patient library or pooled library from several patients) and antibody format (i.e. monoclonal antibody, oligoclonal mixture of antibodies of unknown specificity or patient serum) used for selection, as demonstrated by the varying outcome in similar selections (see section 3.4.1 and 3.4.2).

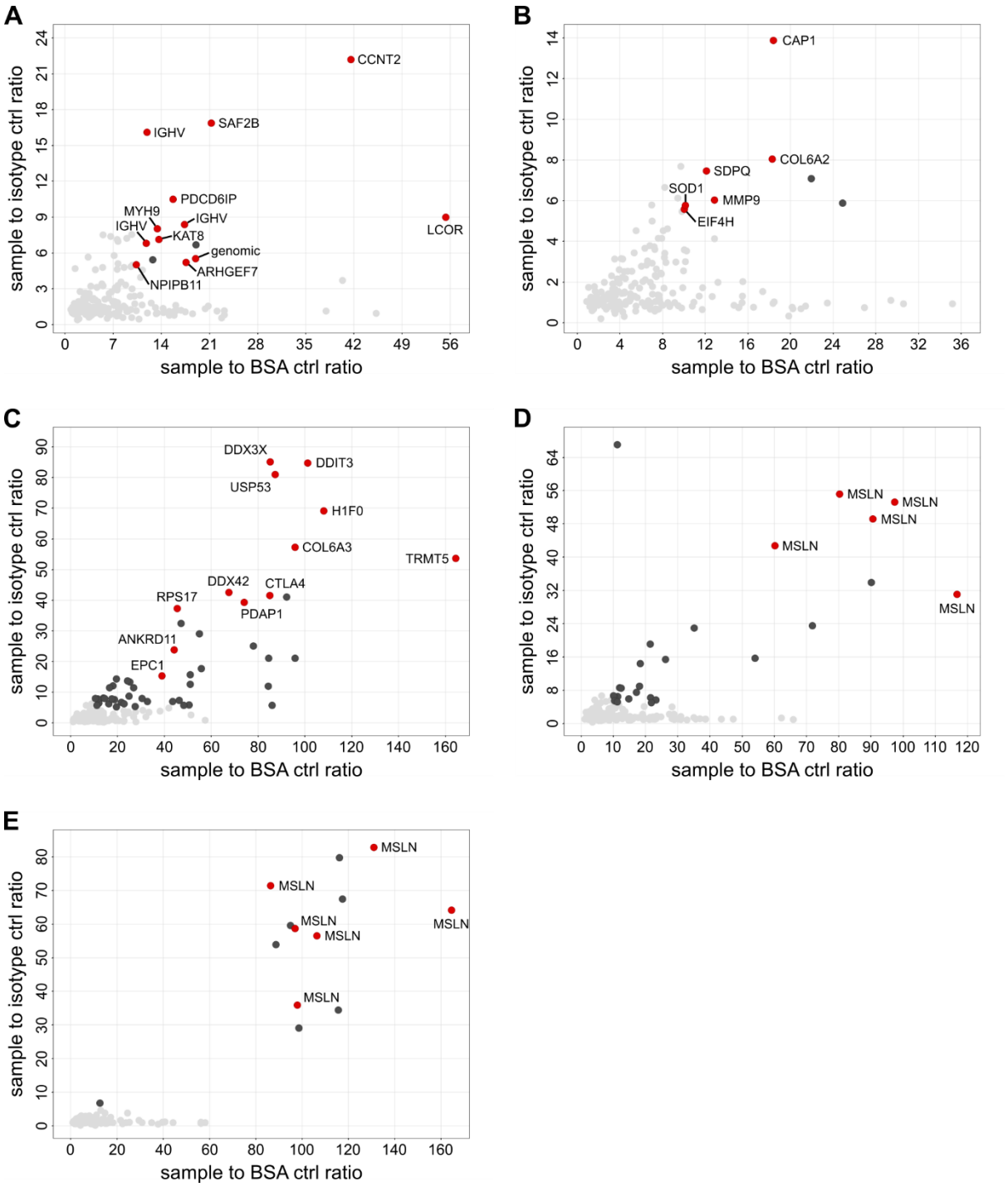


Figure 3-9 – Phage screening ELISA for pannings in 96-well format: *Peptides selected by panning in solution were tested as Hyperphage packaged single clones for their specific binding to different antibodies and antibody mixtures in phage ELISA. Peptides selected from a pooled patient antigen library were tested against (A) an oligoclonal mixture of recombinant scFv-mouse-Fc antibodies from Yuhan0011 TIBC, (B) an oligoclonal mixture of recombinant scFv-mouse-Fc antibodies from Yuhan0012 TIBC (C) Yuhan008 patient serum, (D) monoclonal anti-mesothelin antibody SuW57-D11 in scFv-mouse-Fc format. (E) Peptides selected from HCT116 S4 antigen library were tested against monoclonal anti-mesothelin antibody SuW57-D11 in scFv-mouse-Fc format. Signal to noise ratios to the isotype control antibody GSM238-B7 scFv-mouse-Fc and to BSA are plotted. light grey dots: samples that did not pass the selection criteria; dark grey dots: samples that passed the selection criteria; red dots: samples that passed the selection criteria and were analysed by Sanger sequencing; labels: gene symbols of the best match from NCBI-BLASTn for the selected sequence*

3.3.5 Using NGS as an alternative screening method

Despite multiplexed pannings comprising up to 96 samples were shown to be possible, the next bottleneck of the technology was the need to screen a large number of individual clones from many individual pannings. Usually, binding specificity of single clones is analysed by ELISA. Positive clones are then sequenced to check their identity and enrichment. Performing ELISA of several hundred clones for each of the 96 selections, i.e. of more than 25,000 clones individually, is possible but laborious and time consuming. Moreover, expression or display of antigens during screening may be insufficient to achieve detectable signals in ELISA, resulting in false negative results. Alternatively, by using next generation sequencing (NGS), millions of sequences could be analysed in a single experiment potentially yielding information about all antigen clones enriched after selection. The hypothesis is that despite binding specificity is not analysed in a separate functional assay, the enrichment of sequences during panning could be indicative of the relevance of the encoded polypeptide.

In order to check whether NGS can be used as a surrogate screening method, DNA from either amplified or eluted phage particles from different pannings was analysed. Yuhan007 and Yuhan011 libraries were selected against corresponding oligoclonal pools of scFv-mouse-Fc antibodies and Yuhan008 library was selected against patient serum. For Yuhan007 and Yuhan008, pannings were run in three independent parallel samples for each library. The whole volume of eluted phage particles was collected and analysed after one, two or three rounds of selection, respectively. For Yuhan011, a single panning was performed. Re-amplified phage particles of the second round and eluted phage particles of the third round of selection were sampled for analysis.

The results demonstrated that panning steps altered the distribution of genes in all analysed samples, as expected. Across all samples, individual selection rounds show high variation in gene distribution and gene diversity decreases with increasing selection rounds. In selections carried out with oligoclonal recombinant antibody mixtures, genes that were present in high numbers in the original unselected libraries are still present in abundance after the first selection round but they diminish or are lost in later selection rounds (Figure 3-10 A-D, Figure 3-11 A-D, Figure 3-12 A-C). Meanwhile, genes that show high enrichment after the second and third selection round are not necessarily enriched after the first selection round. This shows the necessity for at least two rounds of selection in this sample type. In the selections carried out with patient serum, the first round already shows strong enrichment of genes that are lowly represented in the unselected library.

To select candidates for further investigation, the 100 genes with the highest read counts were analysed for each selection round. Read counts measured for a gene after selection were divided by the read counts for the same gene in the unselected library to yield an enrichment factor. If a gene was not detected in the unselected library, read counts were set to 1 to generate the enrichment factor. Genes with a read count greater than 20,000 (roughly two times the average read count) and an enrichment factor of more than 200 were regarded as hit candidates.

For Yuhan007 the most abundant cDNA sequences after the second selection round map to WWC2, SHTN1, MTCO2P12, SPRR3 and HSPA8 genes (Supplementary Table 3). Interestingly, while SHTN1 is also the second most frequent gene after the third round, WWC2 and MTCO2P12 are lost and SPRR3 and HSPA8 are lowly represented. Instead PRKDC, EFHD2, CHD8 and PRRC2B sequences were enriched (Supplementary Table 4). This result shows that selections with complex antigen libraries and antibody mixtures are not entirely reproducible. Taking the enrichment factors into account, WWC2, SHTN1, PRKDC, EFHD2, UBE2L6 and CHD8 emerge as hit candidates (Figure 3-10 E).

For Yuhan008 the most enriched sequences after the second panning round map to the PRRC2B gene, followed by EEA1, TOP1MT, CHD8 and UBE2L6 (Supplementary Table 7). After the third round TOP1MT, EEA1, IGHV3-65, CHD8 and UBE2L6 are most abundant (Supplementary Table 8). Due to their respective enrichment factors, all these genes qualify as hit candidates (Figure 3-11). The enrichment of EEA1 and IGHV3-65 is in accordance with results gained from screening ELISA after an identical panning (see section 3.4.2).

For Yuhan011, MMP9 and MYC sequences are the most abundant after the second selection round and after the third round they make up roughly 99 % of all analysed sequences (Figure 3-12 C and D; Supplementary Table 10 and 11). MMP9 enrichment was also confirmed by screening ELISA after an identical panning (see section 3.4.1).

These results show, that NGS enables monitoring de-/enrichment processes during panning and that it may be used as an alternative screening method. In some but not the majority of cases, it yielded the same hits as functional screening by ELISA. However, analysis of only one of the selection rounds might not be enough to draw correct conclusions about the most relevant binding partners.

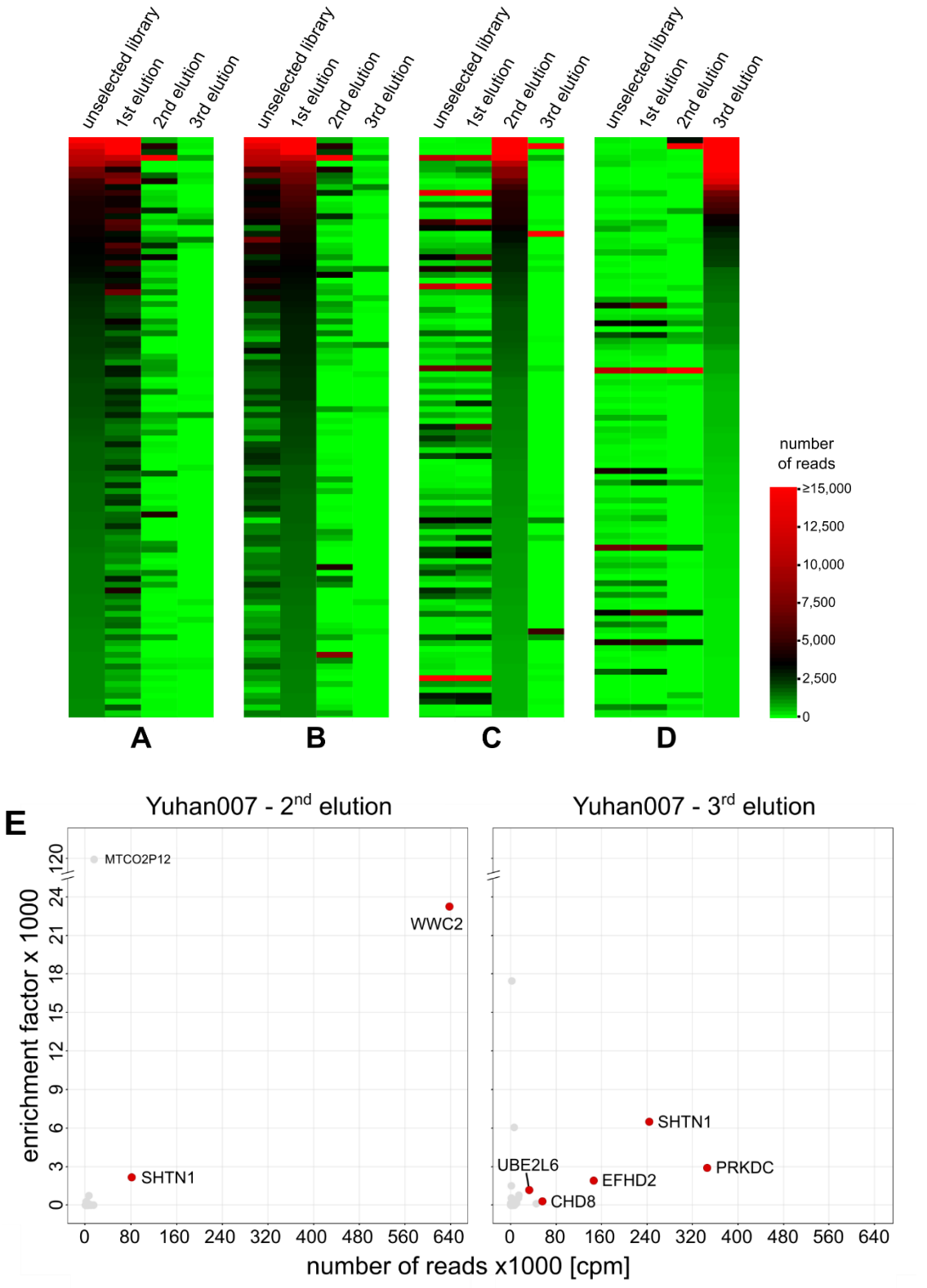


Figure 3-10 – Yuhan007 sequence enrichment: *Three individual pannings were performed on an oligoclonal mixture of recombinant scFv-Fc antibodies derived from TIBC from Yuhan007 using the antigen phage library from the same patient. Eluted phage particles were sampled after round one, two and three, respectively. Unselected library and elution samples were analysed via NGS. Heatmaps show read counts across all samples for the 100 genes with the highest read counts in (A) the unselected library, (B) eluted phage after round one, (C) eluted phage after round two and (D) eluted phage after round three. (E) Read counts of the 100 genes with the highest read count were plotted against the enrichment factor (reads per gene in sample divided by reads per gene in unselected library). Grey dots: genes that did not pass the selection threshold, red dots: genes that passed the selection threshold. Genes are labelled with HGNC gene symbol*

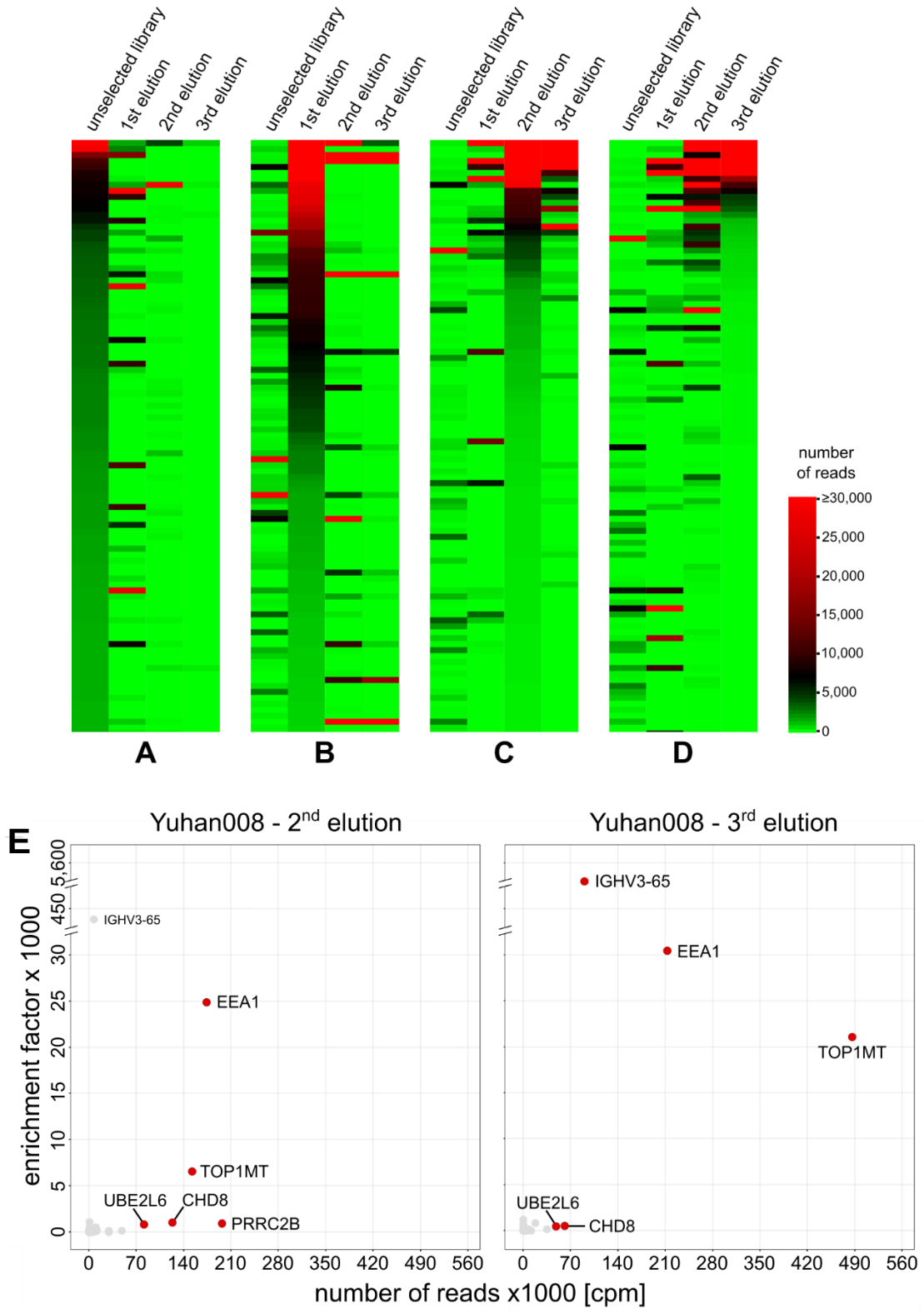


Figure 3-11 – Yuhan008 sequence enrichment: *Three individual pannings were performed on patient serum from Yuhan008 using the antigen phage library from the same patient. Eluted phage particles were sampled after round one, two and three, respectively. Unselected library and elution samples were analysed via NGS. Heatmaps show read counts across all samples for the 100 genes with the highest read counts in (A) the unselected library, (B) eluted phage after round one, (C) eluted phage after round two and (D) eluted phage after round three. (E) Read counts of the 100 genes with the highest read count were plotted against the enrichment factor (reads per gene in sample divided by reads per gene in unselected library). Grey dots: genes that did not pass the selection threshold, red dots: genes that passed the selection threshold. Genes are labelled with HGNC gene symbols*

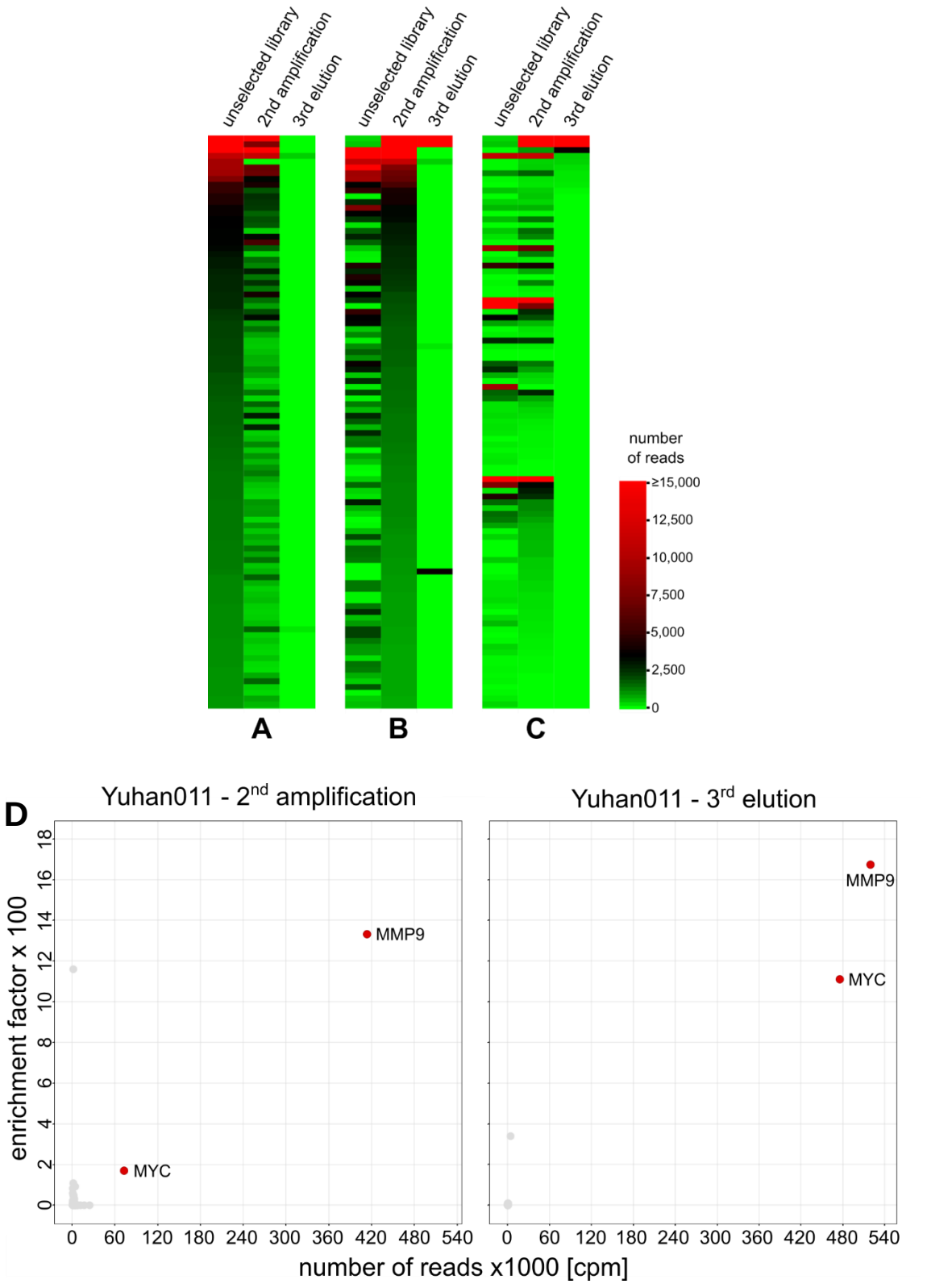


Figure 3-12 – Yuhan011 sequence enrichment: *Panning was performed on an oligoclonal mixture of recombinant scFv-Fc antibodies derived from TIBC from Yuhan011 using the antigen phage library from the same patient. Phage particles were sampled after amplification of selection round two and after elution of selection round three. Unselected library and selected samples were analysed via NGS. Heatmaps show read counts across all samples for the 100 genes with the highest read counts in (A) the unselected library, (B) amplified phage particles after round one and (C) eluted phage after round three. (D) Read counts of the 100 genes with the highest read count were plotted against the enrichment factor (reads per gene in sample* divided by *reads per gene in unselected library). Grey dots: genes that did not pass the selection threshold, red dots: genes that passed the selection threshold. Genes are labelled with HGNC gene symbols*

3.4 Identification of target candidates

In order to identify peptides, which have the potential to be used as therapeutic targets, biomarkers or simply just being tumor associated proteins of interest in the context of head and neck cancer, antigen phage libraries were panned against either oligoclonal mixtures of TIBC derived scFv-Fc antibodies (kindly provided by Melanie Philippi) or against patient sera.

3.4.1 Panning on oligoclonal scFv-Fc

The antigen phage library of Yuhan012 was panned on an oligoclonal mixture of recombinant scFv-Fc antibodies derived from TIBC from the same patient. Screening ELISA was performed with 192 single clones each from the amplified antigen phage particles of the second round and the phage particles eluted after the third round of panning. Clones were regarded as positive when their signal to noise ratio for isotype and BSA control were both greater than 5 (Figure 3-13). A total of 18 clones was chosen for sequencing and 16 of these clones were identified as an MMP9 fragment by BLASTn analysis. The remaining two samples were discarded due to low sequencing quality. When the different selected MMP9 mRNA sequences are aligned, they share a 48 bp overlapping sequence. This suggests that this peptide stretch contains the epitope of the antibody or antibodies that caused the enrichment and selection of the MMP9 peptides.

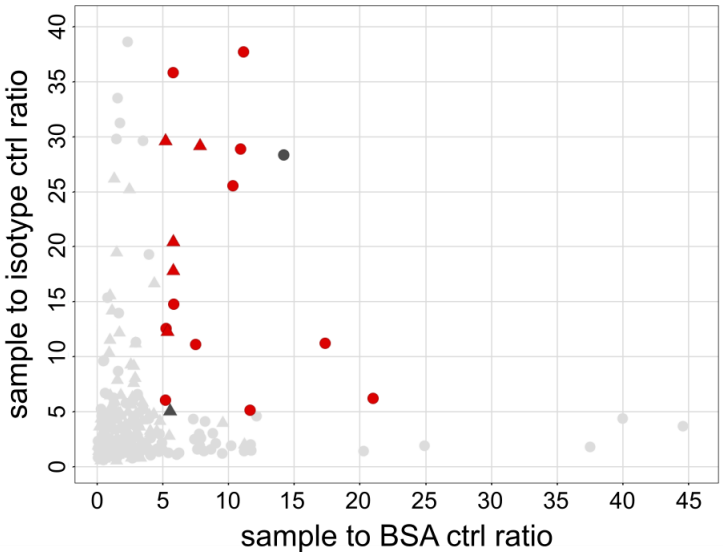


Figure 3-13 – Phage screening ELISA for Yuhan012 panning: *Peptides selected from Yuhan012 antigen library by panning in solution were tested as Hyperphage packaged single clones for their binding to a corresponding recombinant oligoclonal scFv-Fc antibody mixture from TIBC in ELISA. Signal to noise ratios to the isotype control antibody GSM238-B7 scFv-mouse-Fc and to BSA are plotted. light grey symbols: samples that did not pass the selection criteria; dark grey symbols: samples that passed the selection criteria; red symbols: samples that passed the selection criteria and identified as MMP9 mRNA sequences after Sanger sequencing; dots: single clones sampled from the amplified phage particles after the second panning round; triangles: single clones sampled from the eluted phage particles after the third panning round*

3.4.2 Panning on patient serum

The antigen phage library of Yuhan008 was panned on serum of the same patient. Screening ELISA was performed with 192 single clones each from the amplified antigen phage particles of the second round and those eluted after the third round of panning. Clones were regarded as positive, when their signal to noise ratios for isotype and BSA control were both greater than 2. A total of 42 clones was chosen for sequencing and 16 of these clones were identified as an EEA1 fragment by BLASTn analysis. Again, a sequence overlap of the different antigen fragments was found in multiple sequence alignment. The remaining 26 sequences were all identified as an identical mRNA sequence corresponding to an immunoglobulin heavy chain variable region (IGHV3-65). It is noticeable that the IGHV sequences are much more prevalent after the third panning round, whereas the EEA1 sequences have higher occurrence after the second panning round (Figure 3-14).

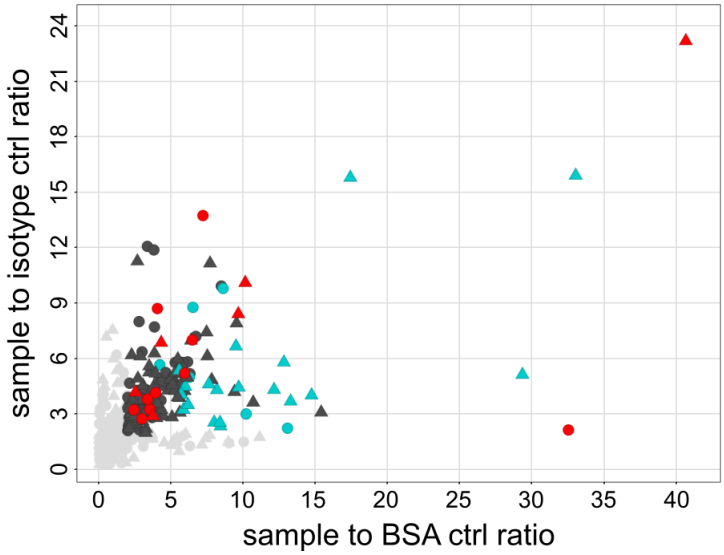


Figure 3-14 – Phage screening ELISA for Yuhan008 serum panning: *Peptides selected from Yuhan008 transcriptome library by panning in solution were tested as Hyperphage packaged single clones for their binding to serum from the same patient in ELISA. Signal to noise ratios to the isotype control antibody GSM238-B7 scFv-mouse-Fc and to BSA are plotted. light grey symbols: samples that did not pass the selection criteria; dark grey symbols: samples that passed the selection criteria; red symbols: samples that passed the selection criteria and identified as EEA1 mRNA sequences after Sanger sequencing; teal symbols: samples that passed the selection criteria and identified as IGHV3-65 mRNA sequences after Sanger sequencing. dots: single clones sampled from the amplified phage particles after the second panning round; triangles: single clones sampled from the eluted phage particles after the third panning round*

4 Discussion

This thesis describes a new approach for the discovery of potential novel biomarkers and therapeutic targets for cancer immunotherapy using cDNA peptide libraries from fresh tumor specimens that are selected against autologous antibody repertoires from tumor infiltrating B cells (TIBC) in addition to autologous patient sera. Phage display peptide libraries have been used extensively in the search for new therapeutic targets, biomarkers and antibody epitopes[29,37,50,51]. Commercially available short peptide libraries enclose large diversities of up to 10^9 unique sequences (for example Ph.D.™-12 from New England Biolabs), yet they are of artificial design and only yield linear epitopes. Other common approaches, like the SEREX technology[52], use cDNA libraries derived from frozen or fixed tissues or cell lines and screen the encoded peptides against patient sera. These libraries represent the natural distribution of expressed genes more closely than synthetic libraries. However, in the case of cell lines, they originate from a model system. Fresh and untreated tissues represent patient biology and disease. The use of fresh unfixed tissue allows separation of different cell types prior to downstream treatments and analysis. These cells are viable and may be propagated if so desired. In this study, this advantage was used to isolate tissue-residing B cells to access the local antibody repertoire of the tumor tissue. Combined with highly diverse cDNA derived protein fragment libraries these antibodies can help to identify new epitopes or proteins that are relevant for individual patient biology or for the general disease context.

4.1 Antigen library quality

The new pTSD3 vector allows efficient unidirectional cloning of cDNA libraries created from fragmented mRNA. Theoretically, this way of cloning should yield roughly 11 % of sequences in a library that contain ORFs that match the natural ORFs of proteins. Unfortunately, comparing pTSD3 insert sequences to ORFs of proteins of the NCBI reference protein database revealed that this is only the case for 2 % of sequences. This is due to a high percentage of sequences containing stop codons. While the presence of stop codons was expected, higher numbers may occur due to non-mRNA contaminations within the library preparation or as a result of RNA fragmentation. If fragmentation does not yield fragments with an equal distribution of all three possible reading frames, the occurrence of correct ORFs may be reduced. However, the rate of correct ORFs is rescued via Hyperphage and increases to 57 % after packaging.

The antigen libraries created in this work consist of 1.7×10^7 - 1×10^8 antigen clones and are comparable in size with other published cDNA libraries[29,53]. According to NGS analysis

the highest number of unique genes represented with more than 10 reads in a single library was 7,169. However, this is based on 4.6 x 10^6 total reads for the according sample (Yuhan007) which covers only 4.6 % of the theoretical maximum of 1 x 10^8 antigen clones. Deeper sequencing runs, i.e. interrogation of a higher number of sequences, would likely reveal greater gene coverage. This is demonstrated by the presence of gene sequences in samples after panning that were not detected in the unselected libraries. The human transcriptome consists of approximately 21,000 genes[54]. Considering that roughly 11,000 to 13,000 of these genes are actively transcribed in a given tissue type[55], the presented antigen library preparation process may be sufficient in order to achieve complete transcriptome coverage. If necessary, input RNA amounts could be increased and several separate library preparations from the same tissue sample could be combined to increase the number of represented genes.

Many of the highly represented genes, for example those coding for Beta-actin (ACTB), Glyceraldehyde-3-Phosphate Dehydrogenase (GAPDH), ribosomal proteins, or histones are housekeeping genes with equal representation in all libraries. Others, for example genes coding for CEA Cell Adhesion Molecule 5 (CEACAM5), Carboxylesterase 1 (CES1) or Small proline rich protein 3 (SPRR3) are more tissue-/cell type-specific and only highly represented in a single libraries. It is not clear whether these differences arise due to the heterogenic nature of the biological samples or due to an amplification bias during antigen library preparation. By applying emulsion PCR for unbiased amplification of antigen fragments prior to library cloning the likelihood of an amplification bias may be reduced. Additionally, this could preserve low-copy transcripts and increase gene coverage[56].

ORF enrichment using Hyperphage resulted in libraries with very high percentages of ORFs. The limitations in regards to insert size after Hyperphage packaging are in accordance with the work of other groups[57]. However, it is unlikely that the size of the secreted polypeptide-pIII fusion protein is the limiting factor, since larger polypeptides can efficiently be displayed on Hyperphage, as exemplified by scFv antibody phage display[42]. Additionally, several mesothelin (MSLN) inserts of more than 400 bp length were found during the SuW57-D11 selections. The antigen libraries were created using randomly fragmented RNA. For the resulting nucleotides the likelihood for the occurrence of a stop codon increases with length. It may be that a large percentage of longer inserts do not yield functional pIII due to an early stop codon and are therefore not packaged into Hyperphage. By cloning a cDNA library that was pre-selected for ORFs into a phagemid vector Verle et al. were able to achieve display of polypeptides of up to 600 aa in a pVI display system[53].

4.2 Technology validation

It was shown that phage ELISA is the superior screening method compared to antigen ELISA regarding signal strength and total number of identified true positive clones. A major advantage of phage ELISA is the signal amplification through detection of major coat protein pVIII via the secondary antibody which increases sensitivity. This facilitates improved detection of antigen phage particles from non-purified culture supernatants. Additionally, multivalent display of peptides on Hyperphage can cause avidity effects that facilitate the detection of weak interaction partners. The increased detection of low affinity interactions is desirable since the aim of the antigen screening is the identification of interaction partners, regardless of binding kinetics.

The enrichment of MSLN sequences using anti-MSLN antibody SuW57-D11 provides proof of principle for the reproducible selection of specifically bound polypeptide antigens from antigen phage libraries derived from fresh patient tissue. The presence of several different selected sequences that overlap partially allows identification of a possible minimal epitope of the used antibody. More detailed knowledge of the epitope is important for development of an antibody drug after target candidate identification.

Epitope structure is a great issue when selecting polypeptides and protein fragments in an in vitro setting. The presented approach employs *E. coli* and phage for the secretion and presentation of antigenic protein fragments. These simple organisms lack the complex folding machinery of mammals and many protein fragments secreted under these conditions will not assume their natural fold. It must be considered that foreign folds and linear epitopes may be the outcome of this approach. However, sequence alignment of all MSLN fragments selected with SuW57-D11 in several experiments revealed a 180 bp i.e. 60 amino acid overlapping region (Supplementary Figure 1). This region is identical in sequence with the published 3D epitope structure of a therapeutic anti-MSLN antibody[58]. This shows that at least some antigen fragments may assume a tertiary structure or even a natural fold when their intrinsic stability is high enough. Nonetheless, moving the presentation of protein fragments from patient tissue derived antigen libraries to a mammalian display system would likely increase the potential to identify structured epitopes.

4.3 Selection of antigen fragments from tumor derived antigen libraries

Using an oligoclonal mixture of antibodies generated from TIBC isolated from head and neck cancer tissue, MMP9 peptide fragments were enriched. MMP9 has strong implications in various cancers[59,60] and several therapeutics targeting MMP9 and other metalloproteinases are currently being developed[61,62], validating the presented approach for the identification of peptides relevant in the context of head and neck cancer. After selection of MMP9 peptides from the antigen libraries, recombinant full-length protein was used to generate MMP9 specific antibodies using recombinant antibody libraries created from TIBC (Melanie Philippi, personal communication). This shows that the selection of MMP9 was based on the antibody repertoire enclosed in the patients' tissues. Interestingly, sequence analysis revealed that the anti-MMP9 antibodies originated from antibody libraries from several different patients, providing evidence that MMP9 peptide selection was not a patient specific artefact (Melanie Philippi, personal communication).

Besides recombinant antibody mixtures derived from patient tissue, autologous patient serum was used to select antigen fragments. Due to reproducible enrichment of several overlapping peptides, EEA1 was proposed as a likely target candidate by this approach. The sequence overlap is indicative of a specific interaction with the encoded peptide epitope as seen for the anti-MSLN antibody. Furthermore, it decreases the likelihood that the enrichment was caused by a technological bias. The high prevalence of a single IGHV sequence enriched during the same selection process is a likely example for a bias in display and amplification. Antibody fragments can easily be expressed in *E. coli* and presented on phage. Such an advantage can lead to the enrichment of a peptide in phage display without specific binding to the selecting interaction partner[63].

EEA1 has been shown to be overexpressed in aggressive prostate tumors[64], but otherwise not much information exists about its implications in cancer. EEA1 is involved in early endosome formation and trafficking and regulates endosome fusion events. High EEA1 expression causes an increase in endosome fusion events, leading to the formation of much larger endosomes[65]. These endosomes can become multivesicular bodies and fuse with the plasma membrane. Thereby they release exosomes, that have been shown to promote tumor growth and metastasis formation[66,67]. Blocking EEA1 may decrease the number of exosomes that are released from cells and alleviate their tumorigenic effects by reducing the formation of large multivesicular bodies.

Besides EEA1, many other intracellular proteins were selected as interaction partners of antibody repertoires from either TIBC or patient serum (data not shown), suggesting that

the patient's immune system reacted to intracellular material. A reason for that may be the unusual exposure of proteins on the cell surface due to apoptotic events in the tumor[68]. An additional upregulation of these proteins in the tumor could further increase the chance of a local immune response. Some of the selected intracellular proteins are associated with actin filaments and microtubules of the cytoskeleton. Changes in the cytoskeleton are crucial for epithelial-mesenchymal transition, cell motility and finally metastasis formation. Actin filaments are especially important for the formation of invadopodia that are needed to cross physiological barriers[69]. The protein cortactin has been shown to be required for initiation and stabilization of invadopodia. SHNT1, which has been selected using oligoclonal mixtures of TIBC derived antibodies (see section 3.3.5.) is normally involved in neuronal polarization and neurite outgrowth. However, it interacts with cortactin to promote signal-force transduction[70]. Its presence in head and neck cancer tissues could suggest an implication of SHTN1 in cancer cell motility. These results imply that there may be a need for more efficient intracellular drug delivery systems to target dysregulated and possibly mutated proteins inside of cancer cells.

As exemplified by the experiment in the 96-well format, the selections using recombinant oligoclonal antibody mixtures or patient sera are not always reproducible. The stoichiometry of this approach is very complex. Millions of different protein fragments interact with up to thousands of different antibodies to reach the interaction with the highest affinity. The resulting binding equilibria may change drastically if one of the complex mixtures of interaction partners is altered. In the mentioned experiment, several antigen libraries were pooled prior to selection. While this did not affect the selection of MSLN fragments with a single antibody, the selections using recombinant oligoclonal antibody mixtures or patient sera failed to yield the same results as when tested against antigen libraries from the same patient.

4.4 Relevance of identified protein fragments

Potential new therapeutic targets and biomarkers identified by application of ORFeome display in combination with TIBC derived antibodies and patient sera originate from an interaction of proteins with patient antibody repertoires. However, a detailed scientific process is needed to validate their relevance in the disease setting. Promising novel targets, by their definition, are not well described. Therefore, finding information on their implications in cancer through literature research is challenging. Harvesting gene expression and proteomics data from publicly available databases can be partially automated and may help to refine the list of candidates but hard evidence can only be

generated through a series of experiments. The first step would be the generation of recombinant antibodies with strong affinity for the suggested epitope. With the presented approach, target binding antibodies are readily available from the TIBC libraries with a high degree of certainty and in a short period of time. These antibodies then need to be tested for their ability to specifically bind to cancerous tissue or specific cells therein. Classic immunohistochemistry on tissue slides is still an acknowledged standard in this regard. Passing this stage, the new target and antibody have the potential to be a new biomarker or diagnostic tool, respectively. Whereas molecular characteristics of the antibody can be modified quite easily, functionality of the antibody is key for its consideration as a therapeutic agent. Besides performing cytotoxicity and internalization assays, the ability to effectively carry cytotoxic payloads should be scrutinized[71]. Such measures have been impeded by time and budget restrictions for the target candidates found in this study but will be considered for future projects. Anyhow, providing evidence for the therapeutic relevance of a target is just a small part of the necessary investigations. A deep understanding of the molecular biology of the target is needed to predict a mode of action and possible side effects of a therapeutic intervention. This includes, among others, the elucidation of its basic function within the cell and regulation of its gene expression and identification of upstream or downstream mediators.

4.5 Implications of next generation sequencing

NGS has previously been used to gain better insight into the selection process and to replace conventional screening methods in phage display[72,73]. It enables a more complete view of the selection outcome and the implementation of high throughput and multi-sample parallelized approaches. In this study, NGS has been used for the same reasons and to test its compatibility with the presented workflow. By comparing sequence frequencies in the unselected and selected samples an enrichment factor can be calculated. Together with the total read count for a gene after selection this factor helps to discriminate likely hit candidates. For this, deep sequencing of the unselected library to detect all enclosed genes is of great importance since missing read count values hinder correct calculation of the enrichment factor. To further refine the selection of specifically enriched gene sequences several negative controls should be included in the panning procedure to rule out those clones that prevail due to propagation advantage or binding to some part of the selection system.

Whereas for small peptide libraries, a single round of panning seems to be enough to reveal candidates via NGS[74], the data from this study suggests that two rounds of selection are needed for libraries with larger peptides that are selected against complex mixtures of antibodies. After the first panning round gene frequencies did not change significantly and genes that were highly enriched after round two and three were sometimes even represented with fewer sequences than in the unselected library. A third panning round may confound the results by enabling unspecific peptides with an amplification advantage to reach very high copy numbers[63]. This is exemplified by the immunoglobulin sequence selected from the Yuhan008 library which replaces EEA1 as the most frequent gene after three panning rounds. When patient sera are used for selection, strong enrichment can already be seen after the first selection round. However, the most highly represented genes are not the same as after two and three rounds of panning, indicating increased reproducibility of selection after performing more than one round of selection

Thus far, NGS has confirmed EEA1 and MMP9 sequences as one of the three most enriched sequences in their according selections. WWC2 and SHTN1 that showed strong enrichment in the NGS data for Yuhan007 have also been selected previously in conventional screening methods (data not shown). By setting cut-off values for total read count and enrichment factor these genes could be discerned from other frequent sequences in the analysis. This shows that NGS can replace conventional screening in a high throughput application. However, for chosen antigen fragments specific binding to an antibody needs to be validated experimentally in any case.

4.6 Outlook

This work provides proof of principle that tumor relevant proteins can be identified using tumor antigen phage display libraries and TIBC-derived antibody repertoires. At this point in time, the application of this technology resides at single patient level and individual findings and insights. Higher sample numbers should be processed for the discovery of proteins that are relevant for many patients or specific patient groups that are defined by disease characteristics. As demonstrated, the method is compatible with high-throughput applications and larger amounts of samples can easily be processed. Furthermore, the technology can be applied to other types of cancer and act as a target discovery platform.

It may be useful for the discovery process to create antibody libraries that keep the natural pairing of heavy and light chain of TIBC derived antibodies. This would likely reduce the complexity of the antibody mixtures used for selection and thereby improve efficiency. Another advancement would be single B cell cloning of antibody sequences. This would offer the possibility to conduct monoclonal selections on the antigen libraries, which has proven to be the more powerful and reproducible approach. Additionally, the identification of many interaction partners of TIBC antibodies will enable conclusions about the local immune reaction. Investigating the antibody repertoires of TIBC populations might give information about why they have beneficial or detrimental influences on clinical outcomes, depending on the biological context[4,75].

NGS is superior to ELISA as a first line screening in terms of throughput and information yield, but ELISA should still be used to confirm binding of the selected target candidates to the antibodies used for selection. By designing and maintaining a bioinformatics pipeline, increasing amounts of data can be analysed and integrated which means that the impact of NGS analyses further improves over time. Besides a better understanding of the antigen selection process this will enable insights into individual patient biology via analysis of the unselected libraries. Since the antigen libraries are generated using methods similar to RNA sequencing library generation, comparative transcriptome analysis should be possible, provided that a sufficient sequencing depth is reached.

Taken together, the information provided in this study may be used to create a target discovery pipeline for the development of antibody drugs.

5 Summary

This study presents an approach to identify proteins with relevance in tumor biology by selecting protein fragments from fresh tumor tissue derived antigen libraries against patient sera and antibody repertoires derived from tumor infiltrating B cells. These proteins have the potential to become new biomarkers or therapeutic targets. Applying this technology, the well-known therapeutic target MMP9 was selected besides newly proposed targets, for example EEA1, WWC2 and SHTN1. While validation protocols for these new targets still need to be set up, the groundwork for a potential target discovery platform is laid out. The method is high throughput applicable and screening efforts can be reduced by implementing sophisticated technologies like NGS. The latter approach also yields the possibility for more in-depth analyses and might allow insights into disease biology as a side product to therapeutic target discovery.

6 Acknowledgements

Many people contribute to a research project, yet in the end, the thesis is attributed to a single person. Knowing that without helping hands, companions, cooperation and supervision a project like this would never truly thrive, I would like to take the opportunity to express my gratefulness to all those people who took part in this scientific journey.

First of all, my deep gratitude goes to my doctoral supervisor and the first examiner of this thesis, **Prof. Dr. Stefan Dübel**. Thank you for the opportunity to participate in such an amazing project that realized my dream of participating in cancer research, for your constant scientific advice and support throughout this work.

Also, I am very grateful for the help of **Prof. Dr. Andreas Gerstner**, the second examiner of this thesis. Thank you for your unclouded enthusiasm, delightful communication and for providing the support and material without which this project would never have gotten this far.

I would like to thank **Prof. Dr. Michael Hust** for sharing his expertise on the pHORF technology and for always being an enthusiastic consult in scientific matters.

My gratitude goes to **Dr. Thomas Schirrmann,** the CEO of Yumab GmbH, who signed me up to this project. Thank you for always keeping the bigger picture in mind and for pushing me from time to time.

Thank you very much, **Dr. Jonas Kügler** and **Dr. André Frenzel** for your guidance and support, for being realistic, for managing expectations and for patiently answering so many questions.

I would like to thank **Dr. Lars Toleikis** for recruiting me to this project and for representing our interests inside and outside of the joined research committee.

My gratitude goes to **Dr. Carolin Sellmann** for the productive discussions and for providing a lot of information and material and access to external research capacities.

I thank **Dr. Thomas Clarke** for performing the next generation sequencing experiments and for providing his expertise and support in data analysis.

Besides supervision and professional scientific support, I received a lot of help in all things large and small and even more personal and moral support from the entire **team of Yumab GmbH and all members from the Department of Biotechnology.** You witnessed me in my best and worst moods. You cheered me up when I was down and added on when times were great. I was always glad to return the favour, if I could. You made working with and alongside you a great and enjoyable experience. Thank you!

Out of this great group of people, I would like to give special mention to two of them: **Wieland Fahr** and **Giulio Russo,** for all the late evenings filled with scientific insights, moral support and philosophic discussions I thank you deeply.

Of course, a lot of encouragement happened behind the scenes. For this, I want to express my deepest gratitude to my **family**, my parents **Marlis** and **Stephan** and my brother **Conrad**. Thank you for believing in me all this time, for supporting my impulses and for enduring my busy schedule that keeps me away from home so much of the time.

In a similar regard I would like to thank my **friends**, that are spread out over so many cities. Thank you for being the colourful, kooky, hilarious and exceptional lot that you are. Thank you for constantly reminding me that I was doing something great, when I myself had forgotten. Thank you for being the balance to all the stressful and troublesome things in life.

And last but surely not least, my profound gratitude goes to **Melanie Philippi**. During the last three years we stood together through thick and thin. You were always there when times were tough, made me laugh and kept me going. Your positive spirit and your good heart are an example to others. I would not want to have made it without you. Thank you!

7 References

1. Hanahan, D. & Weinberg, R. A. Hallmarks of cancer: the next generation. *Cell* **144,** 646–674; 10.1016/j.cell.2011.02.013 (2011).
2. Balkwill, F. R., Capasso, M. & Hagemann, T. The tumor microenvironment at a glance. *Journal of cell science* **125,** 5591–5596; 10.1242/jcs.116392 (2012).
3. Vinay, D. S. *et al.* Immune evasion in cancer: Mechanistic basis and therapeutic strategies. *Seminars in cancer biology* **35 Suppl,** S185-S198; 10.1016/j.semcancer.2015.03.004 (2015).
4. Santoiemma, P. P. & Powell, D. J. Tumor infiltrating lymphocytes in ovarian cancer. *Cancer biology & therapy* **16,** 807–820; 10.1080/15384047.2015.1040960 (2015).
5. Stanton, S. E. & Disis, M. L. Clinical significance of tumor-infiltrating lymphocytes in breast cancer. *Journal for immunotherapy of cancer* **4,** 59; 10.1186/s40425-016-0165-6 (2016).
6. Heimes, A.-S. *et al.* Subtype-specific prognostic impact of different immune signatures in node-negative breast cancer. *Breast cancer research and treatment* **165,** 293–300; 10.1007/s10549-017-4327-0 (2017).
7. Linnebacher, M. & Maletzki, C. Tumor-infiltrating B cells: The ignored players in tumor immunology. *Oncoimmunology* **1,** 1186–1188; 10.4161/onci.20641 (2012).
8. Goc, J., Fridman, W.-H., Sautès-Fridman, C. & Dieu-Nosjean, M.-C. Characteristics of tertiary lymphoid structures in primary cancers. *Oncoimmunology* **2,** e26836; 10.4161/onci.26836 (2013).
9. Aloisi, F. & Pujol-Borrell, R. Lymphoid neogenesis in chronic inflammatory diseases. *Nature reviews. Immunology* **6,** 205–217; 10.1038/nri1786 (2006).
10. Bray, F. *et al.* Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: a cancer journal for clinicians* **68,** 394–424; 10.3322/caac.21492 (2018).
11. Semrau, R. The Role of Radiotherapy in the Definitive and Postoperative Treatment of Advanced Head and Neck Cancer. *Oncology research and treatment* **40,** 347–352; 10.1159/000477128 (2017).
12. Cramer, J. D., Burtness, B. & Ferris, R. L. Immunotherapy for head and neck cancer: Recent advances and future directions. *Oral oncology* **99,** 104460; 10.1016/j.oraloncology.2019.104460 (2019).
13. Falzone, L., Salomone, S. & Libra, M. Evolution of Cancer Pharmacological Treatments at the Turn of the Third Millennium. *Frontiers in pharmacology* **9,** 1300; 10.3389/fphar.2018.01300 (2018).
14. Xin Yu, J., Hubbard-Lucey, V. M. & Tang, J. Immuno-oncology drug development goes global. *Nat Rev Drug Discov*; 10.1038/d41573-019-00167-9 (2019).
15. Jun Tang & Jia Xin Yu. Global Immuno-Oncology Drug Development Pipeline. Cancer Research Institute Clinical Accelerator interactive dashboard. Available at http://www.cancerresearch.org/scientists/immuno-oncology-landscape (2019).

16. Karlitepe, A., Ozalp, O. & Avci, C. B. New approaches for cancer immunotherapy. *Tumour biology : the journal of the International Society for Oncodevelopmental Biology and Medicine* **36,** 4075–4078; 10.1007/s13277-015-3491-2 (2015).

17. Patel, A., Kaufman, H. L. & Disis, M. L. Next generation approaches for tumor vaccination. *Chinese clinical oncology* **6,** 19; 10.21037/cco.2017.02.04 (2017).

18. Scott, A. M., Wolchok, J. D. & Old, L. J. Antibody therapy of cancer. *Nature reviews. Cancer* **12,** 278–287; 10.1038/nrc3236 (2012).

19. Krishnamurthy, A. & Jimeno, A. Bispecific antibodies for cancer therapy: A review. *Pharmacology & therapeutics* **185,** 122–134; 10.1016/j.pharmthera.2017.12.002 (2018).

20. Haanen, J. B. A. G. & Robert, C. Immune Checkpoint Inhibitors. *Progress in tumor research* **42,** 55–66; 10.1159/000437178 (2015).

21. Cohen, J. E. *et al.* Adoptive cell therapy: past, present and future. *Immunotherapy* **9,** 183–196; 10.2217/imt-2016-0112 (2017).

22. Gashaw, I., Ellinghaus, P., Sommer, A. & Asadullah, K. What makes a good drug target? *Drug discovery today* **17 Suppl,** S24-30; 10.1016/j.drudis.2011.12.008 (2012).

23. Failli, M., Paananen, J. & Fortino, V. Prioritizing target-disease associations with novel safety and efficacy scoring methods. *Scientific reports* **9,** 9852; 10.1038/s41598-019-46293-7 (2019).

24. Santos, R. *et al.* A comprehensive map of molecular drug targets. *Nat Rev Drug Discov* **16,** 19–34; 10.1038/nrd.2016.230 (2017).

25. Terstappen, G. C., Schlüpen, C., Raggiaschi, R. & Gaviraghi, G. Target deconvolution strategies in drug discovery. *Nat Rev Drug Discov* **6,** 891–903; 10.1038/nrd2410 (2007).

26. Paananen, J. & Fortino, V. An omics perspective on drug target discovery platforms. *Briefings in bioinformatics*; 10.1093/bib/bbz122 (2019).

27. Liu, B. C.-S., Dijohnson, D. A. & O'Rourke, D. J. Antibody profiling with protein antigen microarrays in early stage cancer. *Expert opinion on medical diagnostics* **6,** 187–196; 10.1517/17530059.2012.672969 (2012).

28. Larkin, S. E. T. *et al.* Detection of candidate biomarkers of prostate cancer progression in serum: a depletion-free 3D LC/MS quantitative proteomics pilot study. *British journal of cancer* **115,** 1078–1086; 10.1038/bjc.2016.291 (2016).

29. Ali Eldib, A. M. *et al.* Immunoscreening of a cDNA library from a lung cancer cell line using autologous patient serum: Identification of XAGE-1b as a dominant antigen and its immunogenicity in lung adenocarcinoma. *International journal of cancer* **108,** 558–563; 10.1002/ijc.11587 (2004).

30. Aronson, J. K. & Ferner, R. E. Biomarkers-A General Review. *Current protocols in pharmacology* **76,** 9.23.1-9.23.17; 10.1002/cpph.19 (2017).

31. Bratkovic, T. Progress in phage display: evolution of the technique and its application. *Cellular and molecular life sciences : CMLS* **67,** 749–767; 10.1007/s00018-009-0192-2 (2010).

32. Breitling, F., Dübel, S., Seehaus, T., Klewinghaus, I. & Little, M. A surface expression vector for antibody screening. *Gene* **104,** 147–153; 10.1016/0378-1119(91)90244-6 (1991).

33. Fühner, V. *et al.* Development of Neutralizing and Non-neutralizing Antibodies Targeting Known and Novel Epitopes of TcdB of Clostridioides difficile. *Frontiers in microbiology* **9,** 2908; 10.3389/fmicb.2018.02908 (2018).

34. Fuchs, M. *et al.* Novel human recombinant antibodies against Mycobacterium tuberculosis antigen 85B. *BMC biotechnology* **14,** 68; 10.1186/1472-6750-14-68 (2014).

35. Fühner, V. *et al.* Epitope Mapping via Phage Display from Single-Gene Libraries. *Methods in molecular biology (Clifton, N.J.)* **1904,** 353–375; 10.1007/978-1-4939-8958-4_17 (2019).

36. Hust, M. *et al.* Enrichment of open reading frames presented on bacteriophage M13 using hyperphage. *BioTechniques* **41,** 335–342; 10.2144/000112225 (2006).

37. Meyer, T. *et al.* Identification of immunogenic proteins and generation of antibodies against Salmonella Typhimurium using phage display. *BMC biotechnology* **12,** 29; 10.1186/1472-6750-12-29 (2012).

38. Naseem, S. *et al.* Phage display-based identification and potential diagnostic application of novel antigens from Mycoplasma mycoides subsp. mycoides small colony type. *Veterinary microbiology* **142,** 285–292; 10.1016/j.vetmic.2009.09.071 (2010).

39. Connor, D. O., Zantow, J., Hust, M., Bier, F. F. & Nickisch-Rosenegk, M. von. Identification of Novel Immunogenic Proteins of Neisseria gonorrhoeae by Phage Display. *PloS one* **11,** e0148986; 10.1371/journal.pone.0148986 (2016).

40. Zantow, J. *et al.* Mining gut microbiome oligopeptides by functional metaproteome display. *Scientific reports* **6,** 34337; 10.1038/srep34337 (2016).

41. Becker, M. *et al.* Application of M13 phage display for identifying immunogenic proteins from tick (Ixodes scapularis) saliva. *BMC biotechnology* **15,** 43; 10.1186/s12896-015-0167-3 (2015).

42. Rondot, S., Koch, J., Breitling, F. & Dübel, S. A helper phage to improve single-chain antibody presentation in phage display. *Nature biotechnology* **19,** 75–78; 10.1038/83567 (2001).

43. Kügler, J. *et al.* Identification of immunogenic polypeptides from a Mycoplasma hyopneumoniae genome library by phage display. *Applied microbiology and biotechnology* **80,** 447–458; 10.1007/s00253-008-1576-1 (2008).

44. Miethe, S. *et al.* Development of Human-Like scFv-Fc Neutralizing Botulinum Neurotoxin E. *PloS one* **10,** e0139905; 10.1371/journal.pone.0139905 (2015).

45. Hust, M., Dübel, S. & Schirrmann, T. Selection of recombinant antibodies from antibody gene libraries. *Methods in molecular biology (Clifton, N.J.)* **408,** 243–255; 10.1007/978-1-59745-547-3_14 (2007).

46. Dübel, S. *et al. Rekombinante Antikörper. Lehrbuch und Kompendium für Studium und Praxis / Stefan Dübel, Frank Breitling, André Frenzel, Thomas Jostock, Andrea L. J.*

Marschall, Thomas Schirrmann, Michael Hust (Springer Spektrum, Berlin, Heidelberg, 2019) - pp63ff.

47. Sandhya, S. *et al.* Length variations amongst protein domain superfamilies and consequences on structure and function. *PloS one* **4,** e4981; 10.1371/journal.pone.0004981 (2009).

48. Xu, D. & Nussinov, R. Favorable domain size in proteins. *Folding & design* **3,** 11–17; 10.1016/S1359-0278(98)00004-2 (1998).

49. Schirrmann, T. & Hust, M. Construction of human antibody gene libraries and selection of antibodies by phage display. *Methods in molecular biology (Clifton, N.J.)* **651,** 177–209; 10.1007/978-1-60761-786-0_11 (2010).

50. Russo, N. *et al.* A novel approach to biomarker discovery in head and neck cancer using an autoantibody signature. *Oncogene* **32,** 5026–5037; 10.1038/onc.2012.532 (2013).

51. Moreira, G. M. S. G., Fühner, V. & Hust, M. Epitope Mapping by Phage Display. *Methods in molecular biology (Clifton, N.J.)* **1701,** 497–518; 10.1007/978-1-4939-7447-4_28 (2018).

52. Türeci, O., Sahin, U. & Pfreundschuh, M. Serological analysis of human tumor antigens: molecular definition and implications. *Molecular medicine today* **3,** 342–349; 10.1016/s1357-4310(97)01081-2 (1997).

53. Somers, V. A. *et al.* A panel of candidate tumor antigens in colorectal cancer revealed by the serological selection of a phage displayed cDNA expression library. *Journal of immunology (Baltimore, Md. : 1950)* **169,** 2772–2780; 10.4049/jimmunol.169.5.2772 (2002).

54. The ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489,** 57–74; 10.1038/nature11247 (2012).

55. Ramsköld, D., Wang, E. T., Burge, C. B. & Sandberg, R. An abundance of ubiquitously expressed genes revealed by tissue transcriptome sequence data. *PLoS computational biology* **5,** e1000598; 10.1371/journal.pcbi.1000598 (2009).

56. Williams, R. *et al.* Amplification of complex gene libraries by emulsion PCR. *Nature methods* **3,** 545–550; 10.1038/nmeth896 (2006).

57. Faix, P. H. *et al.* Phage display of cDNA libraries: enrichment of cDNA expression using open reading frame selection. *BioTechniques* **36,** 1018-22, 1024, 1026-9; 10.2144/04366RR03 (2004).

58. Ma, J., Tang, W. K., Esser, L., Pastan, I. & Di Xia. Recognition of mesothelin by the therapeutic antibody MORAb-009: structural and mechanistic insights. *The Journal of biological chemistry* **287,** 33123–33131; 10.1074/jbc.M112.381756 (2012).

59. Huang, H. Matrix Metalloproteinase-9 (MMP-9) as a Cancer Biomarker and MMP-9 Biosensors: Recent Advances. *Sensors (Basel, Switzerland)* **18**; 10.3390/s18103249 (2018).

60. Wu, Q.-W. *et al.* Expression and clinical significance of matrix metalloproteinase-9 in lymphatic invasiveness and metastasis of breast cancer. *PloS one* **9,** e97804; 10.1371/journal.pone.0097804 (2014).

61. Koslawsky, D. *et al.* A bi-specific inhibitor targeting IL-17A and MMP-9 reduces invasion and motility in MDA-MB-231 cells. *Oncotarget* **9,** 28500–28513; 10.18632/oncotarget.25526 (2018).

62. Juric, V. *et al.* MMP-9 inhibition promotes anti-tumor immunity through disruption of biochemical and physical barriers to T-cell trafficking to tumors. *PloS one* **13,** e0207255; 10.1371/journal.pone.0207255 (2018).

63. Vodnik, M., Zager, U., Strukelj, B. & Lunder, M. Phage display: selecting straws instead of a needle from a haystack. *Molecules (Basel, Switzerland)* **16,** 790–817; 10.3390/molecules16010790 (2011).

64. Johnson, I. R. D. *et al.* Endosomal gene expression: a new indicator for prostate cancer patient prognosis? *Oncotarget* **6,** 37919–37929; 10.18632/oncotarget.6114 (2015).

65. Ramanathan, H. N., Zhang, G. & Ye, Y. Monoubiquitination of EEA1 regulates endosome fusion and trafficking. *Cell & bioscience* **3,** 24; 10.1186/2045-3701-3-24 (2013).

66. Kim, D. H. *et al.* Exosomal PD-L1 promotes tumor growth through immune escape in non-small cell lung cancer. *Experimental & molecular medicine* **51,** 94; 10.1038/s12276-019-0295-2 (2019).

67. Peinado, H. *et al.* Melanoma exosomes educate bone marrow progenitor cells toward a pro-metastatic phenotype through MET. *Nature medicine* **18,** 883–891; 10.1038/nm.2753 (2012).

68. Hansen, M. H., Nielsen, H. V. & Ditzel, H. J. Translocation of an intracellular antigen to the surface of medullary breast cancer cells early in apoptosis allows for an antigen-driven antibody response elicited by tumor-infiltrating B cells. *Journal of immunology (Baltimore, Md. : 1950)* **169,** 2701–2711; 10.4049/jimmunol.169.5.2701 (2002).

69. Fife, C. M., McCarroll, J. A. & Kavallaris, M. Movers and shakers: cell cytoskeleton in cancer metastasis. *British journal of pharmacology* **171,** 5507–5523; 10.1111/bph.12704 (2014).

70. Kubo, Y. *et al.* Shootin1-cortactin interaction mediates signal-force transduction for axon outgrowth. *The Journal of cell biology* **210,** 663–676; 10.1083/jcb.201505011 (2015).

71. Wang, X., An, Z., Luo, W., Xia, N. & Zhao, Q. Molecular and functional analysis of monoclonal antibodies in support of biologics development. *Protein & cell* **9,** 74–85; 10.1007/s13238-017-0447-x (2018).

72. Dias-Neto, E. *et al.* Next-generation phage display: integrating and comparing available molecular tools to enable cost-effective high-throughput analysis. *PloS one* **4,** e8338; 10.1371/journal.pone.0008338 (2009).

73. Ravn, U. *et al.* By-passing in vitro screening--next generation sequencing technologies applied to antibody display and in silico candidate selection. *Nucleic acids research* **38,** e193; 10.1093/nar/gkq789 (2010).

74. Hoen, P. A. C. 't *et al.* Phage display screening without repetitious selection rounds. *Analytical biochemistry* **421,** 622–631; 10.1016/j.ab.2011.11.005 (2012).

75. Spector, M. E. *et al.* Prognostic Value of Tumor-Infiltrating Lymphocytes in Head and Neck Squamous Cell Carcinoma. *JAMA otolaryngology-- head & neck surgery*; 10.1001/jamaoto.2019.2427 (2019).

8 Appendix

8.1 Supplementary data

Supplementary Table 1 – Yuhan007 unselected library NGS data: Three individual pannings were performed on an oligoclonal mixture of recombinant scFv-Fc antibodies derived from TIBC from Yuhan007 using the antigen phage library from the same patient. Eluted phage particles were sampled after round one, two and three, respectively. Unselected library and elution samples were analysed via NGS. Read counts across all samples for the twenty genes with the highest read counts in the unselected library are shown.

HGNC gene symbol	**Yuhan007 unselected library reads [cpm]**	Yuhan007 1st elution reads [cpm]	Yuhan007 2nd elution reads [cpm]	Yuhan007 3rd elution reads [cpm]
MT-ND2	28389.00	26750.17	675.83	63.63
GAPDH	13376.28	15961.04	4511.44	119.31
KRT6A	11176.96	15148.62	2231.49	0.00
SPRR3	10501.40	11279.68	14870.43	757.21
TMSB4X	10193.27	8729.36	0.00	0.00
EEF1A1	7264.28	4131.65	8.31	0.00
KRT17	6898.29	7122.05	1537.04	230.66
MT-CO3	5060.19	8061.56	4123.28	127.26
RPLP0	4777.76	3215.78	0.00	0.00
KRT5	4610.00	4951.33	250.05	3.18
FTH1	4285.91	3831.75	362.84	3.18
ANXA1	4106.91	3864.46	274.54	1.59
RNA18S5	4104.36	4784.29	2573.12	157.49
PABPC1	4102.54	3035.57	179.24	386.56
ACTB	4052.64	5988.23	493.70	1253.28
MT-ND4	3792.75	5614.77	0.87	0.00
CTTN	3770.88	3153.21	493.99	1.59
KRT15	3594.47	3430.46	951.26	1221.72
KRT19	3432.08	5862.25	2597.60	173.39
KRT16	3417.87	4540.29	541.20	1.59

Supplementary Table 2 – Yuhan007 1st elution sample NGS data: Three individual pannings were performed on an oligoclonal mixture of recombinant scFv-Fc antibodies derived from TIBC from Yuhan007 using the antigen phage library from the same patient. Eluted phage particles were sampled after round one, two and three, respectively. Unselected library and elution samples were analysed via NGS. Read counts across all samples for the twenty genes with the highest read counts in the 1st elution sample are shown. Enrichment factors are calculated by dividing the read count for a gene in the eluted sample by the read count in the unselected library.

HGNC gene symbol	Yuhan007 unselected library reads [cpm]	**Yuhan007 1st elution reads [cpm]**	Yuhan007 2nd elution reads [cpm]	Yuhan007 3rd elution reads [cpm]	Yuhan007 1st elution enrichment factor
MT-ND2	28389.00	26750.17	675.83	63.63	0.94
GAPDH	13376.28	15961.04	4511.44	119.31	1.19
KRT6A	11176.96	15148.62	2231.49	0.00	1.36
SPRR3	10501.40	11279.68	14870.43	757.21	1.07
TMSB4X	10193.27	8729.36	0.00	0.00	0.86
MT-CO3	5060.19	8061.56	4123.28	127.26	1.59
KRT17	6898.29	7122.05	1537.04	230.66	1.03
MUC4	2698.65	6869.50	1206.56	0.00	2.55
ACTB	4052.64	5988.23	493.70	1253.28	1.48
KRT19	3432.08	5862.25	2597.60	173.39	1.71
MT-ND4	3792.75	5614.77	0.87	0.00	1.48
MT-ATP6	3294.67	5268.70	194.05	0.00	1.60
KRT5	4610.00	4951.33	250.05	3.18	1.07
RNA18S5	4104.36	4784.29	2573.12	157.49	1.17
KRT16	3417.87	4540.29	541.20	1.59	1.33
CD68	1313.08	4286.78	48.96	0.00	3.26
MUC21	2737.18	4141.69	226.45	0.00	1.51
EEF1A1	7264.28	4131.65	8.31	0.00	0.57
ANXA1	4106.91	3864.46	274.54	1.59	0.94
FTH1	4285.91	3831.75	362.84	3.18	0.89

Supplementary Table 3 – Yuhan007 2nd elution sample NGS data: Three individual pannings were performed on an oligoclonal mixture of recombinant scFv-Fc antibodies derived from TIBC from Yuhan007 using the antigen phage library from the same patient. Eluted phage particles were sampled after round one, two and three, respectively. Unselected library and elution samples were analysed via NGS. Read counts across all samples for the twenty genes with the highest read counts in the 2nd elution sample are shown. Enrichment factors are calculated by dividing the read count for a gene in the eluted sample by the read count in the unselected library.

HGNC gene symbol	Yuhan007 unselected library reads [cpm]	Yuhan007 1st elution reads [cpm]	**Yuhan007 2nd elution reads [cpm]**	Yuhan007 3rd elution reads [cpm]	Yuhan007 2nd elution enrichment factor
WWC2	27.37	80.99	637029.62	0.00	23272.63
SHTN1	37.40	37.92	81270.45	243226.50	2173.10
MTCO2P12	0.13	0.00	15251.60	0.00	119795.29
SPRR3	10501.40	11279.68	14870.43	757.21	1.42
HSPA8	760.77	521.62	10454.23	35.00	13.74
MT-ATP8	707.80	1298.67	8169.67	0.00	11.54
STX6	258.69	541.55	8090.06	0.00	31.27
SPRYD7	8.51	0.00	6214.85	0.00	729.95
IFITM2	155.01	72.41	5183.82	0.00	33.44
GAPDH	13376.28	15961.04	4511.44	119.31	0.34
EEF1A1P5	33.79	0.00	4427.68	0.00	131.04
NCL	1523.79	1542.38	4339.24	0.00	2.85
USP11	105.83	39.53	4228.20	0.00	39.95
ZNF512B	124.05	24.43	4206.34	0.00	33.91
MT-CO3	5060.19	8061.56	4123.28	127.26	0.81
MYH9	3377.86	3308.12	3707.98	6.36	1.10
PRKDC	118.96	43.71	3153.66	345548.36	26.51
RPS20	283.67	149.45	3140.55	0.00	11.07
SPTBN1	445.28	533.52	2851.15	0.00	6.40
HNRNPA2B1	849.02	631.54	2691.15	0.00	3.17

Supplementary Table 4 – Yuhan007 3rd elution sample NGS data: *Three individual pannings were performed on an oligoclonal mixture of recombinant scFv-Fc antibodies derived from TIBC from Yuhan007 using the antigen phage library from the same patient. Eluted phage particles were sampled after round one, two and three, respectively. Unselected library and elution samples were analysed via NGS. Read counts across all samples for the twenty genes with the highest read counts in the 3rd elution sample are shown. Enrichment factors are calculated by dividing the read count for a gene in the eluted sample by the read count in the unselected library.*

HGNC gene symbol	Yuhan007 unselected library reads [cpm]	Yuhan007 1st elution reads [cpm]	Yuhan007 2nd elution reads [cpm]	**Yuhan007 3rd elution reads [cpm]**	Yuhan007 3rd elution enrichment factor
PRKDC	118.96	43.71	3153.66	345548.36	2904.77
SHTN1	37.40	37.92	81270.45	243226.50	6503.65
EFHD2	75.43	179.34	0.00	145602.13	1930.21
CHD8	183.73	150.41	56.83	55452.86	301.82
PRRC2B	335.31	52.71	20.11	44661.08	133.19
UBE2L6	27.21	0.64	3.50	32292.75	1186.65
UACA	18.54	0.00	179.24	14211.99	766.56
HOOK3	24.11	6.11	0.00	13119.13	544.14
XAF1	65.57	0.00	20.98	9834.18	149.99
LMNB2	260.36	404.96	0.00	7048.73	27.07
IGHV1-24	1.00	0.00	1.75	6079.95	6079.95
EEA1	12.81	8.03	0.00	5539.08	432.37
RALBP1	94.05	132.09	782.51	5187.52	55.16
CEP164	25.70	6.11	0.00	3779.68	147.06
HSPB1	408.20	604.22	0.00	3673.10	9.00
CLDN4	768.31	303.40	0.00	2950.89	3.84
ATF6	78.59	11.89	4.37	2712.27	34.51
PARP10	169.80	0.96	143.39	2429.11	14.31
YTHDF1	124.93	132.09	6.99	2386.16	19.10
TRIM29	482.04	538.34	0.00	2285.95	4.74

Supplementary Table 5 – Yuhan008 unselected library NGS data: Three individual pannings were performed on patient serum from Yuhan008 using the antigen phage library from the same patient. Eluted phage particles were sampled after round one, two and three, respectively. Unselected library and elution samples were analysed via NGS. Read counts across all samples for the twenty genes with the highest read counts in the unselected library are shown.

HGNC gene symbol	**Yuhan008 unselected library reads [cpm]**	Yuhan008 1st elution reads [cpm]	Yuhan008 2nd elution reads [cpm]	Yuhan008 3rd elution reads [cpm]
IGHG1	29639.85	1540.79	4211.78	755.82
TMSB4X	28523.99	2492.23	0.93	1.55
IGKC	15390.61	15753.67	0.00	0.00
HSPA1A	10995.89	15.83	53.22	0.00
HSPA1B	10164.28	0.00	0.00	4.65
IGHA1	8462.14	0.00	0.00	0.00
CD74	8310.08	208.60	1.87	0.00
AL928742.1 (Linc RNA)	7680.61	1477.50	29045.21	225.04
ACTB	7596.42	50355.58	0.93	20.15
HSPA1A	7429.24	9950.23	0.00	0.00
EEF1A1	7164.20	0.00	0.00	57.34
HBA1	6870.07	4.24	91.50	0.00
GAPDH	6174.60	0.32	85.90	159.61
FTH1	5880.37	9072.72	94.30	23.24
CD68	4804.90	0.00	0.00	0.00
RNA18S5	4326.85	1479.82	0.00	1.55
IGHG4	4218.92	0.32	1483.93	1.60
HSPA1B	3753.62	0.00	0.00	0.00
MYH9	3648.22	948.99	304.38	3.10
HLA-B	3479.10	0.00	0.00	0.00

Supplementary Table 6 – Yuhan008 1st elution sample NGS data: Three individual pannings were performed on patient serum from Yuhan008 using the antigen phage library from the same patient. Eluted phage particles were sampled after round one, two and three, respectively. Unselected library and elution samples were analysed via NGS. Read counts across all samples for the twenty genes with the highest read counts in the unselected library are shown. Read counts across all samples for the twenty genes with the highest read counts in the 1st elution sample are shown. Enrichment factors are calculated by dividing the read count for a gene in the eluted sample by the read count in the unselected library.

HGNC gene symbol	Yuhan008 unselected library reads [cpm]	**Yuhan008 1st elution reads [cpm]**	Yuhan008 2nd elution reads [cpm]	Yuhan008 3rd elution reads [cpm]	**Yuhan008 1st elution enrichment factor**
XAF1	215.00	97209.35	29449.65	3526.97	452.14
KRT16	640.13	69783.18	0.00	7.75	109.01
CHD8	116.54	69780.61	123178.71	60843.37	598.76
PRRC2B	212.88	65914.52	196604.53	34614.27	309.63
ACTB	7596.42	50355.58	0.93	20.15	6.63
IDE	11.65	41698.40	0.00	4.65	3578.00
DGUOK	35.46	31895.56	0.00	1.55	899.56
JUNB	3208.55	27934.13	17.74	0.00	8.71
IVNS1ABP	1455.00	27337.31	0.93	0.00	18.79
ANKRD36C	16.16	26736.30	98.97	0.00	1654.15
RSRP1	152.76	26154.29	0.00	0.00	171.21
PPIAP22	58.54	24244.65	0.00	0.53	414.12
FOXP3	423.43	22220.86	0.00	0.00	52.48
ELOA	53.54	19799.13	0.00	1.55	369.81
DDX39A	183.50	19482.69	115.78	18.60	106.17
IGKC	15390.61	15753.67	0.00	0.00	1.02
CGN	5.58	15315.23	0.00	0.00	2744.74
SBNO2	465.77	14357.55	553.68	0.00	30.83
SCYL1	277.37	12040.76	1046.67	125.52	43.41
TNFAIP3	1952.24	11783.87	0.00	0.00	6.04

Supplementary Table 7 – Yuhan008 2nd elution sample NGS data: Three individual pannings were performed on patient serum from Yuhan008 using the antigen phage library from the same patient. Eluted phage particles were sampled after round one, two and three, respectively. Unselected library and elution samples were analysed via NGS. Read counts across all samples for the twenty genes with the highest read counts in the unselected library are shown. Read counts across all samples for the twenty genes with the highest read counts in the 2nd elution sample are shown. Enrichment factors are calculated by dividing the read count for a gene in the eluted sample by the read count in the unselected library.

HGNC gene symbol	Yuhan008 unselected library reads [cpm]	Yuhan008 1st elution reads [cpm]	**Yuhan008 2nd elution reads [cpm]**	Yuhan008 3rd elution reads [cpm]	**Yuhan008 2nd elution enrichment factor**
PRRC2B	212.88	65914.52	196604.53	34614.27	923.54
EEA1	6.99	647.04	174022.19	212885.42	24887.13
TOP1MT	23.10	0.00	152225.05	486378.16	6590.89
CHD8	116.54	69780.61	123178.71	60843.37	1056.96
UBE2L6	100.01	10091.58	80961.67	48616.74	809.51
HSPB1	325.04	266.54	48751.04	10951.28	149.98
XAF1	215.00	97209.35	29449.65	3526.97	136.97
AL928742.1 (Linc RNA)	7680.61	1477.50	29045.21	225.04	3.78
RALBP1	46.62	0.32	11619.84	7830.31	249.27
IGHV1-24	34.44	0.97	11031.61	4066.25	320.30
TNKS1BP1	139.35	0.00	10987.73	959.22	78.85
CEP164	20.98	757.46	10406.94	17868.93	496.10
DCAF15	46.40	849.52	10033.49	653.95	216.22
PARP10	477.18	5007.97	8785.14	175.11	18.41
IGHV3-65	0.02	27.03	7406.66	90174.42	436934.81
DCAF6	61.38	6665.81	6157.72	4563.68	100.32
ATF6	34.75	1121.22	5027.95	912.74	144.68
ERO1B	35.26	527.93	4397.71	97.63	124.72
IGHG1	29639.85	1540.79	4211.78	755.82	0.14
YTHDF1	226.58	2717.25	4123.20	511.38	18.20

Supplementary Table 8 – Yuhan008 3rd elution sample NGS data: Three individual pannings were performed on patient serum from Yuhan008 using the antigen phage library from the same patient. Eluted phage particles were sampled after round one, two and three, respectively. Unselected library and elution samples were analysed via NGS. Read counts across all samples for the twenty genes with the highest read counts in the unselected library are shown. Read counts across all samples for the twenty genes with the highest read counts in the 3rd elution sample are shown. Enrichment factors are calculated by dividing the read count for a gene in the eluted sample by the read count in the unselected library.

HGNC gene symbol	Yuhan008 unselected library reads [cpm]	Yuhan008 1st elution reads [cpm]	Yuhan008 2nd elution reads [cpm]	**Yuhan008 3rd elution reads [cpm]**	**Yuhan008 3rd elution enrichment factor**
TOP1MT	23.10	0.00	152225.05	486378.16	21058.71
EEA1	6.99	647.04	174022.19	212885.42	30445.01
IGHV3-65	0.02	27.03	7406.66	90174.42	5319582.49
CHD8	116.54	69780.61	123178.71	60843.37	522.08
UBE2L6	100.01	10091.58	80961.67	48616.74	486.10
PRRC2B	212.88	65914.52	196604.53	34614.27	162.60
CEP164	20.98	757.46	10406.94	17868.93	851.82
HSPB1	325.04	266.54	48751.04	10951.28	33.69
RALBP1	46.62	0.32	11619.84	7830.31	167.97
DCAF6	61.38	6665.81	6157.72	4563.68	74.35
IGHV1-24	34.44	0.97	11031.61	4066.25	118.06
XAF1	215.00	97209.35	29449.65	3526.97	16.40
NAA10	29.37	0.00	1610.62	2215.98	75.45
TRNAU1AP	10.10	0.00	944.90	1094.04	108.32
TNKS1BP1	139.35	0.00	10987.73	959.22	6.88
ATF6	34.75	1121.22	5027.95	912.74	26.26
IGHG1	29639.85	1540.79	4211.78	755.82	0.03
DCAF15	46.40	849.52	10033.49	653.95	14.09
LINC00514	49.58	0.00	2085.88	587.31	11.85
ANXA4	53.48	0.00	327.73	533.07	9.97

***Supplementary Table 9* – Yuhan011 unselected library NGS data:** Panning was performed on an oligoclonal mixture of recombinant scFv-Fc antibodies derived from TIBC from Yuhan011 using the antigen phage library from the same patient. Phage particles were sampled after amplification of selection round two and after elution of selection round three. Unselected library and selected samples were analysed via NGS. Read counts across all samples for the twenty genes with the highest read counts in the unselected library are shown.

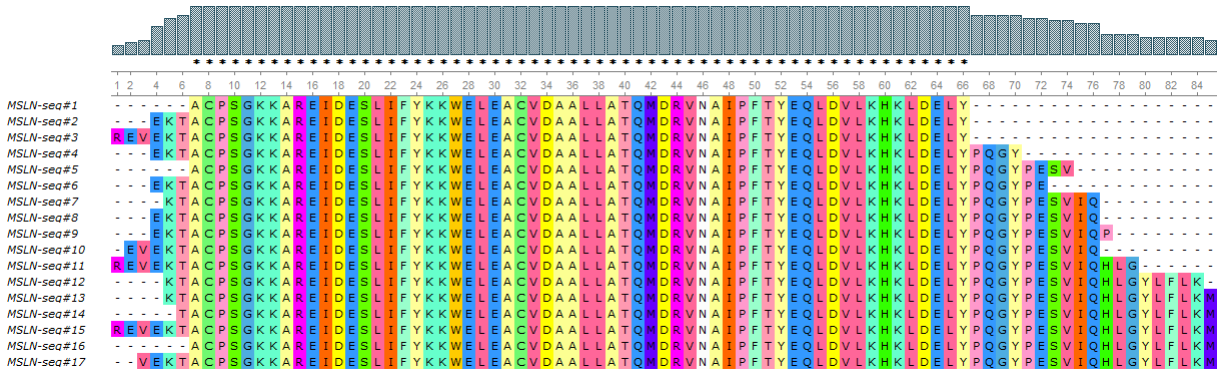
HGNC gene symbol	**Yuhan011 unselected library reads [cpm]**	Yuhan011 2nd amplification reads [cpm]	Yuhan011 3rd elution reads [cpm]
KRT17	38981.93	23845.59	0.94
SPP1	24289.03	7757.36	2.83
KRT5	16286.92	16677.48	2.83
GAPDH	11102.94	11222.52	353.27
HSPA1A	9545.94	0.00	1.43
HSPA1A	9545.94	6508.33	0.00
KRT6A	9325.96	6911.77	6.14
ACTB	7673.30	3335.24	0.94
ITGB4	4955.00	4136.38	3.78
IGHG1	4901.87	1877.51	0.00
FLNA	4363.99	2516.12	0.94
IGKC	4323.74	2335.02	0.00
PKM	3726.00	2276.18	0.47
H3F3B	3679.00	1592.35	0.00
MYH9	3553.95	1958.62	2.83
IGHG4	3506.35	1797.10	0.00
SFN	3471.00	350.02	0.00
ALDH3A1	3401.00	3325.24	0.47
LAMB3	3391.00	5573.54	0.47
ENO1	3311.80	1712.58	0.00

Supplementary Table 10 – Yuhan011 2nd amplification sample NGS data: Panning was performed on an oligoclonal mixture of recombinant scFv-Fc antibodies derived from TIBC from Yuhan011 using the antigen phage library from the same patient. Phage particles were sampled after amplification of selection round two and after elution of selection round three. Unselected library and selected samples were analysed via NGS. Read counts across all samples for the twenty genes with the highest read counts in the amplified sample are shown. Enrichment factors are calculated by dividing the read count for a gene in the eluted sample by the read count in the unselected library.

HGNC gene symbol	Yuhan011 unselected library reads [cpm]	**Yuhan011 2nd amplification reads [cpm]**	Yuhan011 3rd elution reads [cpm]	Yuhan011 2nd amplification enrichment factor
MMP9	309.98	413440.74	519102.57	1333.77
MYC	428.00	72782.61	475374.10	170.05
KRT17	38981.93	23845.59	0.94	0.61
KRT5	16286.92	16677.48	2.83	1.02
GAPDH	11102.94	11222.52	353.27	1.01
SPP1	24289.03	7757.36	2.83	0.32
KRT6A	9325.96	6911.77	6.14	0.74
HSPA1A	9545.94	6508.33	0.00	0.68
LAMB3	3391.00	5573.54	0.47	1.64
ITGB4	4955.00	4136.38	3.78	0.83
ASL	43.55	4074.49	0.00	93.56
EEF1A1	2595.79	4045.24	0.00	1.56
ACTB	7673.30	3335.24	0.94	0.43
ALDH3A1	3401.00	3325.24	0.47	0.98
ACTG1	2295.76	3270.10	0.00	1.42
YBX1	205.97	2866.12	0.94	13.92
MT-CO3	1677.92	2859.35	0.47	1.70
LMNA	2742.00	2779.40	0.00	1.01
TPT1	1610.15	2729.17	1.42	1.69
FADS2	462.87	2607.43	0.00	5.63

Supplementary Table 11 – Yuhan011 3rd amplification sample NGS data: Panning was performed on an oligoclonal mixture of recombinant scFv-Fc antibodies derived from TIBC from Yuhan011 using the antigen phage library from the same patient. Phage particles were sampled after amplification of selection round two and after elution of selection round three. Unselected library and selected samples were analysed via NGS. Read counts across all samples for the twenty genes with the highest read counts in the 3rd elution sample are shown. Enrichment factors are calculated by dividing the read count for a gene in the eluted sample by the read count in the unselected library.

HGNC gene symbol	Yuhan011 unselected library reads [cpm]	Yuhan011 2nd amplification reads [cpm]	**Yuhan011 3rd elution reads [cpm]**	Yuhan011 3rd elution enrichment factor
MMP9	309.98	413440.74	519102.57	1674.63
MYC	428.00	72782.61	475374.10	1110.69
UBA7	11.00	922.75	3751.80	341.07
GAPDH	11102.94	11222.52	353.27	0.03
SLBP	38.00	343.94	345.71	9.10
TRIP10	56.00	383.51	215.36	3.85
COL1A1	1065.00	1654.76	170.49	0.16
IGLV1-40	34.85	213.57	167.18	4.80
CCDC149	14.00	156.75	146.88	10.49
AHNAK	481.00	353.58	32.59	0.07
CNST	20.00	421.04	25.50	1.28
LGALS3	328.64	308.93	22.67	0.07
TALDO1	591.00	647.29	19.36	0.03
ACADM	25.00	57.32	15.11	0.60
SLC2A1	550.00	1374.23	14.64	0.03
PEAR1	6.00	62.40	11.33	1.89
ZFP36L2	439.00	1512.72	8.03	0.02
MLF2	198.00	1112.47	7.08	0.04
IGHV3-65	1.00	2.44	6.61	6.61
KRT6A	9325.96	6911.77	6.14	0.00



Supplementary Figure 1 – MSLN sequence alignment: amino acid sequences of MSLN fragments selected from after different pannings with anti-MSLN antibody SuW57-D11 were aligned using ClustalW. Sequences longer than position 85 in the alignment view were cropped to improve visibility.